REVIEW ARTICLE

Using multiscale norms to quantify mixing and transport

Jean-Luc Thiffeault

Department of Mathematics, University of Wisconsin – Madison, Madison, WI, USA

E-mail: jeanluc@math.wisc.edu

Abstract. Mixing is relevant to many areas of science and engineering, including the pharmaceutical and food industries, oceanography, atmospheric sciences, and civil engineering. In all these situations one goal is to quantify and often then to improve the degree of homogenisation of a substance being stirred, referred to as a passive scalar or tracer. A classical measure of mixing is the variance of the concentration of the scalar, which can be related to the L^2 norm of the concentration field. Recently other norms have been used to quantify mixing, in particular the *mix-norm* as well as negative Sobolev norms. These norms have the advantage that unlike variance they decay even in the absence of diffusion, and their decay corresponds to the flow being mixing in the sense of ergodic theory. General Sobolev norms weigh scalar gradients differently, and are known as *multiscale norms* for mixing. We review the applications of such norms to mixing and transport, and show how they can be used to optimise the stirring and mixing of a decaying passive scalar. We then review recent work on the less-studied case of a continuously-replenished scalar field — the source-sink problem. In that case the flows that optimally reduce the norms are associated with transport rather than mixing: they push sources onto sinks, and vice versa.

PACS numbers: 47.51.+a, 47.52.+j

1. Introduction

One of the most vexing questions about fluid mixing is how to measure it. People typically know it when they see it, but specific applications require customised measures. For example, a measure might be too fine-grained for some applications that don't require thorough mixing. Some measures, such as residence time distributions, are designed for open-flow situations where fluid particles are only mixed for a certain amount of time. Others, such as the rigorous definition of a mixing flow in ergodic theory, are better suited to an idealised mathematical treatment. Finally, one of the main points of this review is that measures used to quantify mixing in the initial-value decaying problem must be interpreted very differently when sources and sinks are present.

One of the earliest attempts to quantify mixing was by the chemical engineer and bomb disposal officer Peter V. Danckwerts [26]. Danckwerts realised that scale was an important consideration; he identified the large-scale breakup of fluid into clumps and the subsequent homogenisation at small scales due to diffusion as separate processes [25]:

> The breaking-up and the interdiffusion are, in the case of liquids, largely independent processes which produce distinguishable results. The former reduces the size of the clumps, while the latter tends to obliterate differences of concentration between neighbouring regions of the mixture. It therefore seems desirable to use two quantities to describe the degree of mixing — namely the *scale of segregation* and the *intensity of segregation*.

Two other pioneers are the oceanographers Carl Eckart [32] and Pierre Welander [92], who also identified the complementary roles of mechanical stirring and diffusion. Following Eckart, modern parlance refers to these two stages as *stirring* and *mixing*, the distinguishing feature being that stirring is a mechanical action, whilst mixing is diffusion-driven or the result of coarse-graining. Welander in particular was emphatic about the important role of stirring in creating *filaments*, which can subsequently be smoothed by diffusion. It was then Batchelor [12] who identified the length scale of the filaments, at which stirring and diffusion achieve a balance, now known as the Batchelor scale.

Let us pin a mathematical meaning on the ideas above. Danckwert's scale of segregation is a correlation length of the concentration of a mixture. His intensity of segregation is a normalised variance of the concentration. It is through the variance that the connection to norms first appears, in this case as the L^2-norm of the concentration field. If $\theta(\boldsymbol{x}, t)$ is the concentration of a passive scalar — such as temperature, dye, or salt — then the variance is

$$\operatorname{Var}\theta = \|\theta\|^2 - \langle\theta\rangle^2, \tag{1.1}$$

where

$$\|\theta\|^2 = \frac{1}{|\Omega|}\int_\Omega \theta^2 \,\mathrm{d}\Omega, \qquad \langle\theta\rangle = \frac{1}{|\Omega|}\int_\Omega \theta \,\mathrm{d}\Omega \tag{1.2}$$

are the L^2-norm and mean value of θ, respectively, with Ω the spatial domain and $|\Omega|$ its volume (or area in two dimensions).

Why is the variance a good measure of mixing quality? A first answer is that it measures fluctuations from the mean, and a mixed state is exactly one where the concentration is equal to the mean — i.e., it is uniform. But there is a second, more intimate reason why variance is important: from the classical advection-diffusion equation for an incompressible velocity field $\boldsymbol{u}(\boldsymbol{x}, t)$ and diffusion coefficient κ,

$$\frac{\partial\theta}{\partial t} + \boldsymbol{u}\cdot\nabla\theta = \kappa\Delta\theta, \qquad \nabla\cdot\boldsymbol{u} = 0, \tag{1.3}$$

we find that the concentration variance is monotonically driven to zero in the absence of sources (see section 3). Indeed, after a few integrations by parts and assuming boundary

conditions that conserve the total amount of θ (no-flux or periodic, see section 2), we find the L^2-norm and mean (1.2) obey

$$\frac{d}{dt}\langle\theta\rangle = 0, \qquad \frac{d}{dt}\|\theta\|^2 = -2\kappa\|\nabla\theta\|^2, \tag{1.4}$$

so that

$$\frac{d}{dt}\mathrm{Var}\,\theta = -2\kappa\|\nabla\theta\|^2. \tag{1.5}$$

Observe that the right-hand side is negative-definite unless $\theta = \text{const.}$, i.e., the concentration is uniform. Hence, the advection–diffusion equation says that the concentration is driven to a uniform state with $\mathrm{Var}\,\theta = 0$, at a rate dictated by the product $2\kappa\|\nabla\theta\|_2^2$. Determining this rate is an important aspect of the *scalar decay* or *initial value* problem, where we have some initial concentration field $\theta(\boldsymbol{x},t) = \theta_0(\boldsymbol{x})$ and want to know how fast it is mixed by a velocity field $\boldsymbol{u}(\boldsymbol{x},t)$. There is a vast literature focused on determining and estimating this decay rate. Note that the advecting velocity field does not appear directly in (1.5): its role is to increase gradients of concentration.

Thus, monitoring variance is a simple way of quantifying the effectiveness of a mixing process. It has, as mentioned above, the dual advantages of being intuitive and of being mathematically sound. In what situations, then, is the variance a less-than-ideal measure of mixing quality? The answer is: when the diffusivity κ is very small and the stirring process is very effective. The typical situation in that case is that the decay term $2\kappa\|\nabla\theta\|_2^2$ becomes *independent* of κ. The physical picture is that the stirring sharpens gradients of θ until they are large enough that diffusion easily smooths them out. A balance between advection and diffusion is then reached, and the variance decays at its optimal rate.

In theory, this is fine; in practice, it is disastrous. If we are trying to optimise the mixing process, it means we have to keep track of scales down to lengths of order $\kappa^{1/2}$, which can be very small. Our simulations are limited by the small quantity κ, and yet the decay rate is independent of κ. However, if we try and avoid this by setting $\kappa = 0$ in (1.3), we find that variance is exactly conserved. The limit is singular: taking κ to zero is completely different from setting $\kappa = 0$.

This problem is similar to what happens for turbulence: letting the viscosity tend to zero in the Navier–Stokes equations does not necessarily recover solutions to Euler's equations. However, here the problem is less severe, since the concentration θ does not feed back on the flow. Hence, it is sensible to solve the pure advection equation

$$\frac{\partial\theta}{\partial t} + \boldsymbol{u}\cdot\nabla\theta = 0 \tag{1.6}$$

and try to extract some measure of mixing from θ; it's just not possible to use variance as a measure. The key is to use a different norm that downplays the role of small scales. For example, a simple choice is $\|\nabla^{-1}\theta\|_2^2$. The operator ∇^{-1} will be defined more carefully later, but for now it suffices to understand that it 'smooths out' θ, so that small-scale variations are not detected. (We make sure the operator is well-defined by restricting to functions with vanishing mean, $\langle\theta\rangle = 0$.)

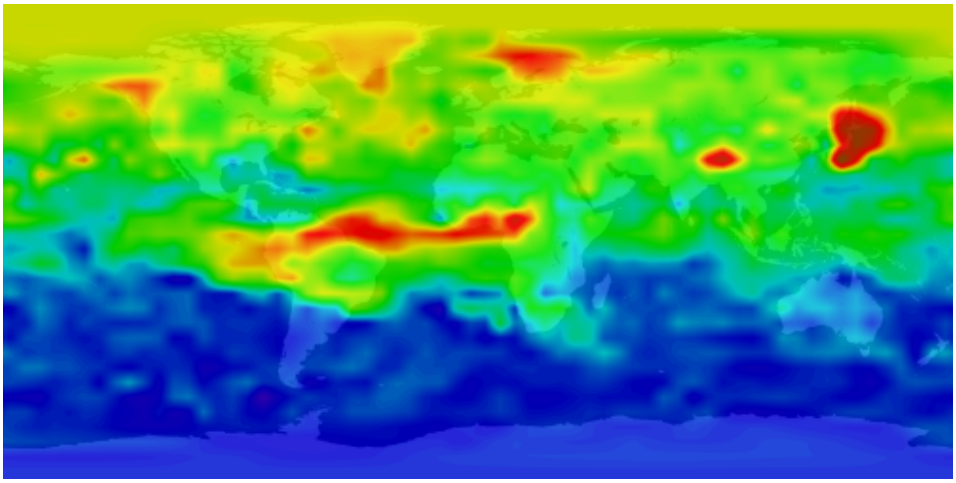


Figure 1. Satellite observations showing concentrations of carbon monoxide (CO) in the atmosphere. Red corresponds to high levels of CO (450 parts per billion) and blue to low levels (50 ppb). Note the immense clouds due to grassland and forest fires in Africa and South America (photo from NASA/NCAR/CSA).

So far our discussion has been of equation (1.3) when the total quantity of scalar is conserved, i.e., $\langle\theta\rangle = \text{const}$. This is the case most often discussed in the literature, and in recent years much work has gone into understanding the factors affecting the decay rate of the concentration and its variance. There is also a vast literature on the case where there are sources and sinks, either within the fluid domain or at its boundaries. In geophysical and industrial situations this is often more relevant. Figure 1 shows a satellite image of the concentration of carbon monoxide in the atmosphere. Observe that there are several large and many small sources, distributed in a complex manner throughout the globe. This begs the question: what do we mean by well-mixed in a case as in figure 1? Unmixed features are persistent, since they are replenished by an inhomogeneous source. But we can still use a norm-measure such as the variance: it will not tend towards zero with time, but a smaller value of variance still indicates that the passive scalar is getting mixed to some degree. A stirring flow that is effective at mixing will thus presumably tend to reduce the variance of the concentration field.

It has become apparent in recent years that the quality of mixing we obtain will depend strongly on the source-sink distribution, in addition to depending on the stirring flow itself. Let us give a simple but extreme example of this, discussed by Plasting & Young [64] and Shaw *et al* [71]. Consider a two-dimensional biperiodic square domain where sources and sinks are present, as in figure 2(a). The equation to be solved is the analogue of (1.3) with a body source term:

$$\frac{\partial\theta}{\partial t} + \boldsymbol{u}\cdot\nabla\theta = \kappa\Delta\theta + s(\boldsymbol{x},t), \tag{1.7}$$

with periodic boundary conditions. Here we choose a simple source-sink distribution, $s(\boldsymbol{x}) = \sin(2\pi x/L)$. If we fix the kinetic energy of the flow, $\frac{1}{2}\left\|\rho\boldsymbol{u}\right\|^2$,‡ what is the

‡ We consider incompressible flows with constant density throughout, so that fixing the kinetic energy

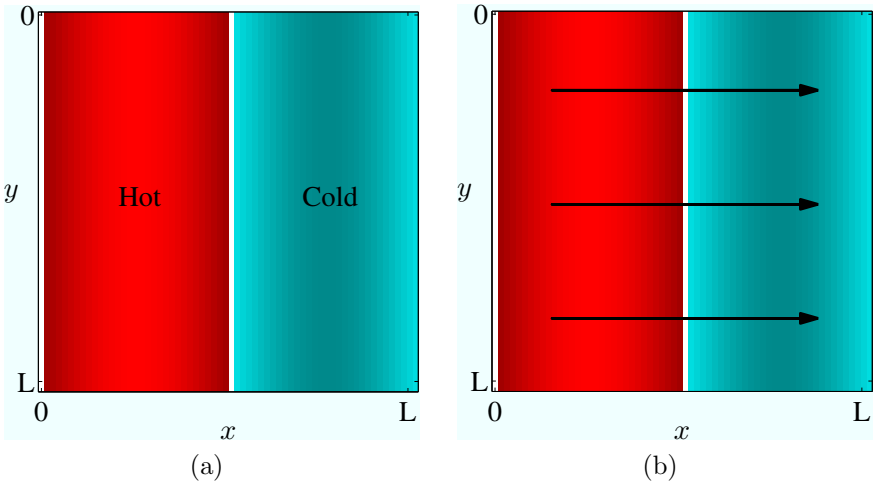


Figure 2. (a) A source-sink distribution in a two-dimensional periodic square domain. (b) Sketch of the velocity field that most effectively reduces the concentration variance of the source.

incompressible stirring velocity field that will most effectively reduce the variance? The answer sketched in figure 2(b) is somewhat surprising: the most effective stirring consists of a uniform (constant) flow. This flow carries the hot fluid onto the cold sink, and cold onto hot, thereby reducing the concentration variance as much as possible.

The reason this is surprising is that 'common wisdom' in mixing assumes that the best stirring is either turbulent or exhibits chaotic trajectories [4, 59]. Such complex behaviour increases concentration gradients and thus allows diffusion to act more effectively. However, this particular source-sink configuration is best mixed not by creating small scales, but rather by transporting fluid appropriately. One might not call this 'mixing' in the strict sense, but it is dependent on diffusion, as (1.7) does not converge to a steady-state without it.

Mixing a scalar field whose fluctuations are constantly replenished by spatially inhomogeneous sources and sinks is a problem with a long history. Townsend [86, 87] was concerned with the effect of turbulence and molecular diffusion on a line source of temperature — a heated filament. The spatial localisation of the source, imposed by experimental constraints, enhanced the role of molecular diffusivity. Durbin [31] and Drummond [30] introduced stochastic particle models to turbulence modelling, and these allowed more detailed studies of the effect of the source on diffusion. Sawford & Hunt [68] pointed out that small sources, such as heated filaments, lead to an explicit dependence of the variance on molecular diffusivity. Many refinements to these models followed, see for instance [2, 3, 15, 85] and the review by Sawford [67]. Chertkov *et al* [16–20] and Balkovsky & Fouxon [8] treated the case of a random, statistically-steady source.

We now give a brief outline of the review. In section 1.3 we recall some basic properties of the advection-diffusion equation, to complement the earlier material in

is equivalent to fixing the L^2-norm $\|\boldsymbol{u}\|^2$. For that reason, we shall often refer to $\|\boldsymbol{u}\|^2$ itself as the kinetic energy.

this introduction. Section 3 is devoted to a review of Sobolev norms. The rest of the review is divided into two parts. Part I discusses norms as measures of mixing for the freely-decaying passive scalar. Section 4 connects norms with the concept of mixing in the sense of ergodic theory. In section 5 we show how negative Sobolev norms can be used to optimise flows to achieve rapid mixing.

Part II, which forms the bulk of the paper, is devoted to advection and diffusion in the presence of sources and sinks. In section 6 we introduce *mixing efficiencies*, measures of mixing based on norms. We give some upper bounds on these efficiencies in section 4. In section 8 we investigate the dependence of efficiencies on functional features of the source-sink distribution. We derive mixing efficiencies from a homogenisation theory approach in section 9. In section 10 we discuss optimisation of mixing efficiencies. Finally, we offer some closing comments in section 11.

2. The advection–diffusion equation

The main equation discussed in this review is the advection–diffusion equation for a passive scalar with concentration field $\theta(\boldsymbol{x}, t)$,

$$\frac{\partial \theta}{\partial t} + \boldsymbol{u} \cdot \nabla \theta = \kappa \Delta \theta + s(\boldsymbol{x}, t), \qquad \nabla \cdot \boldsymbol{u} = 0. \tag{2.1}$$

Following the usual route, we will assume that the domain of interest Ω is a periodic square box with dimension d, with spatial period L in each direction. Of course, everything discussed in this paper can be repeated for a more general closed domain with no-flux boundary conditions, but this adds little to the discussion. In particular, by using a periodic domain we can use Fourier series expansions, which makes many calculations explicit.

The mean $\langle \theta \rangle$ satisfies

$$\frac{d}{dt} \langle \theta \rangle = \langle s \rangle \,, \tag{2.2}$$

with solution

$$\langle \theta \rangle(t) = \langle \theta \rangle(0) + \int_0^t \langle s \rangle(t') \, \mathrm{d}t'. \tag{2.3}$$

Thus, if we replace θ by a new variable

$$\theta'(\boldsymbol{x}, t) = \theta(\boldsymbol{x}, t) - \langle \theta \rangle(t) \tag{2.4}$$

then θ' obeys the modified equation

$$\frac{\partial \theta'}{\partial t} + \boldsymbol{u} \cdot \nabla \theta' = \kappa \Delta \theta' + s'(\boldsymbol{x}, t), \qquad s'(\boldsymbol{x}, t) = s(\boldsymbol{x}, t) - \frac{d}{dt} \langle \theta \rangle \,. \tag{2.5}$$

The new concentration field θ' and source s' have spatial mean zero. For the remainder of this paper, we drop the primes and assume without loss of generality that both θ and s have zero spatial mean.

3. Norms

3.1. Definitions and basic properties

In this section we introduce the measures of mixing we'll be using for the rest of the paper. The Sobolev norm we use for the space $H^q(\Omega)$ is

$$\|f\|_{H^q} = \left(\frac{1}{|\Omega|} \int_\Omega |(1 - L^2\Delta)^{q/2} f|^2 \, \mathrm{d}\Omega \right)^{1/2}, \tag{3.1}$$

or in terms of Fourier series,

$$\|f\|_{H^q} = \Big(\sum_{\boldsymbol{k}} (1 + k^2 L^2)^q \, |\hat{f}_{\boldsymbol{k}}|^2 \Big)^{1/2}, \tag{3.2}$$

where $k \coloneqq |\boldsymbol{k}|$ and $\hat{f}_{\boldsymbol{k}}$ are the Fourier coefficients. In this review we will prefer to use the seminorm on the homogeneous space $\dot{H}^q(\Omega)$ (note the dot over H),

$$\|f\|_{\dot{H}^q} = \left(\frac{1}{|\Omega|} \int_\Omega |(-\Delta)^{q/2} f|^2 \, \mathrm{d}\Omega \right)^{1/2} = \left\| (-\Delta)^{q/2} f \right\|, \tag{3.3}$$

or in terms of Fourier series

$$\|f\|_{\dot{H}^q} = \Big(\sum_{\boldsymbol{k}} k^{2q} \, |\hat{f}_{\boldsymbol{k}}|^2 \Big)^{1/2}. \tag{3.4}$$

If we take a function f with Fourier coefficients behaving asymptotically as $|\hat{f}_{\boldsymbol{k}}| \sim k^p$, $k \gg 1$, then the norms (3.1)–(3.4) converge (exist) for $q + p < -d/2$, where d is the dimension of space.

The norm (3.3) has a more intimate connection with solutions of the advection–diffusion equation than (3.1), as will be described later (section 3.2). Note that the manner in which we have defined (3.1) and (3.3) allows for q positive, negative, or even fractional. In section 2 we showed that we could assume $\langle \theta \rangle = 0$, so we restrict attention to functions f with mean zero. In that case (3.3) becomes a true norm since $\|f\|_{\dot{H}^q} = 0$ if and only if f is zero.

In fact it does not matter much which of the two norms (3.1) or (3.3) we use, since they are equivalent for zero-mean functions: by Poincaré's inequality, we have $\left\| (-\Delta)^{q/2} f \right\| \geq (2\pi/L)^q \, \|f\|$ for $q \geq 0$, so that

$$\|f\|_{\dot{H}^q} \leq L^{-q} \, \|f\|_{H^q} \leq (1 + (2\pi)^{-2})^{q/2} \, \|f\|_{\dot{H}^q}, \qquad q \geq 0, \tag{3.5}$$

for all zero-mean functions f. For $q < 0$ Poincaré's inequality is reversed, so we have

$$(1 + (2\pi)^{-2})^{q/2} \, \|f\|_{\dot{H}^q} \leq L^{-q} \, \|f\|_{H^q} \leq \|f\|_{\dot{H}^q}, \qquad q < 0. \tag{3.6}$$

Equivalence means that if one of the two equivalent norms goes to zero, then the other must as well, and they must do so at the same rate [54, 55].

For the mathematically minded, we can give a rigorous definition of what kind of functions live in the negative Sobolev space $\dot{H}^q$, with $q < 0$, given that we understand the space $\dot{H}^{-q}$. The space $\dot{H}^q$ is defined as the space dual to $\dot{H}^{-q}$ with respect to the standard pairing

$$\langle f\,,\,g\rangle = \frac{1}{|\Omega|}\int_\Omega f(\boldsymbol{x})\,g(\boldsymbol{x})\,\mathrm{d}\Omega, \tag{3.7}$$

with the dual norm

$$\|f\|_{\dot{H}^{-q*}} = \sup_{g\in\dot{H}^{-q}} \frac{\langle f\,,\,g\rangle}{\|g\|_{\dot{H}^{-q}}}\,. \tag{3.8}$$

The norm (3.3) is equal to the dual norm (3.8). To show this, first observe that by using the Cauchy–Schwarz inequality,

$$\frac{\langle f\,,\,g\rangle}{\|g\|_{\dot{H}^{-q}}} = \frac{\left\langle (-\Delta)^{q/2} f\,,\,(-\Delta)^{-q/2} g\right\rangle}{\|g\|_{\dot{H}^{-q}}} \le \|f\|_{\dot{H}^q} \tag{3.9}$$

independent of g, so $\|f\|_{\dot{H}^{-q*}} \le \|f\|_{\dot{H}^q}$. Now let $g = (-\Delta)^{2q} f$:

$$\frac{\langle f\,,\,g\rangle}{\|g\|_{\dot{H}^{-q}}} = \frac{\langle f\,,\,(-\Delta)^{2q} f\rangle}{\|(-\Delta)^{2q} f\|_{\dot{H}^{-q}}} = \frac{\|(-\Delta)^q f\|^2}{\|(-\Delta)^q f\|} = \|f\|_{\dot{H}^q}\,. \tag{3.10}$$

Since the dual norm is defined as a sup over g, we have $\|f\|_{\dot{H}^{-q*}} \ge \|f\|_{\dot{H}^q}$. We conclude that $\|f\|_{\dot{H}^{-q*}} = \|f\|_{\dot{H}^q}$. The same argument can also be used to show $\|f\|_{H^{-q*}} = \|f\|_{H^q}$ for the inhomogeneous spaces.

3.2. Evolution in time

To get a feel for what these norms are telling us about mixing, it is helpful to examine how they evolve in time. That is, given that θ obeys (2.1), what is $d\,\|\theta\|_{\dot{H}^q}/dt$? We start from

$$\tfrac{1}{2}\frac{d}{dt}\,\|\theta\|^2_{\dot{H}^q} = \langle (-\Delta)^q\theta\,\partial_t\theta\rangle \tag{3.11}$$

where $\langle\cdot\rangle$ denotes an average over the periodic domain Ω. Inserting (2.1) for $\partial_t\theta$,

$$\begin{aligned}\tfrac{1}{2}\frac{d}{dt}\,\|\theta\|^2_{\dot{H}^q} &= \langle (-\Delta)^q\theta\,(-\boldsymbol{u}\cdot\nabla\theta + \kappa\Delta\theta + s)\rangle \\ &= -\langle (-\Delta)^q\theta\,\boldsymbol{u}\cdot\nabla\theta\rangle - \kappa\left\|(-\Delta)^{(q+1)/2}\theta\right\|^2 + \langle (-\Delta)^q\theta\,s\rangle\,.\end{aligned}$$

The case $q = 0$ gives the evolution of the variance, for which the velocity term on the right integrates away:

$$\tfrac{1}{2}\frac{d}{dt}\,\|\theta\|^2_{\dot{H}^0} = -\kappa\,\|\nabla\theta\|^2 + \langle\theta\,s\rangle\,. \tag{3.12}$$

There are two other cases that give particularly nice equations. The case $q = 1$ gives the evolution of scalar concentration gradients:

$$\tfrac{1}{2}\frac{d}{dt}\,\|\theta\|^2_{\dot{H}^1} = -\langle\nabla\theta\cdot\nabla\boldsymbol{u}\cdot\nabla\theta\rangle - \kappa\,\|\Delta\theta\|^2 - \langle\Delta\theta\,s\rangle\,. \tag{3.13}$$

The first term on the right is the familiar 'stretching' term, which says that gradients are increased or decreased proportionally to their alignment with the principal axes of the rate-of-strain tensor. A direction of positive strain will decrease gradients, whilst a direction of negative strain will increase gradients.

The final case of interest to us is $q = -1$, for which

$$\tfrac{1}{2}\frac{d}{dt}\left\|\theta\right\|_{\dot{H}^{-1}}^2 = \left\langle \nabla^{-1}\theta \cdot \nabla \boldsymbol{u} \cdot \nabla^{-1}\theta \right\rangle - \kappa \left\|\theta\right\|^2 - \left\langle \Delta^{-1}\theta\, s \right\rangle. \tag{3.14}$$

Here we interpret ∇^{-1} via its action on Fourier modes

$$(\nabla^{-1}\theta)_{\boldsymbol{k}} = -\frac{i\boldsymbol{k}}{k^2}\,\hat{\theta}_{\boldsymbol{k}}, \tag{3.15}$$

so that $\nabla \cdot \nabla^{-1}\theta = \theta$. (Recall that we are restricting to functions with vanishing mean.)

Compare the first term on the right-hand side of (3.14) to the same term for (3.13): velocity gradients have the opposite effect on $\frac{d}{dt}\left\|\theta\right\|_{\dot{H}^{-1}}$ as they do on $\frac{d}{dt}\left\|\theta\right\|_{\dot{H}^{1}}$. This is intuitively clear: the creation of concentration gradients will tend to make θ very filamented. We will see in section 4 that this will cause it to converge weakly to zero, and that this implies that any negative Sobolev norm must go to zero. Another argument that the two norms should evolve with opposite trends arises from

$$\left\|\theta\right\|^2 = -\left\langle \nabla\theta \cdot \nabla^{-1}\theta \right\rangle \le \left\|\theta\right\|_{\dot{H}^{1}} \left\|\theta\right\|_{\dot{H}^{-1}}. \tag{3.16}$$

Since the variance $\left\|\theta\right\|^2$ is conserved when $\kappa = s = 0$, inequality (3.16) implies that if the norm $\left\|\theta\right\|_{\dot{H}^{-1}}$ converges to zero, then $\left\|\theta\right\|_{\dot{H}^{1}}$ must diverge.

Part I: The decaying problem

4. Mixing in the sense of ergodic theory

A divergence-free velocity field $\boldsymbol{u}(\boldsymbol{x}, t)$ generates a time-dependent function $\theta(\boldsymbol{x}, t)$ via the advection equation,

$$\frac{\partial\theta}{\partial t} + \boldsymbol{u}\cdot\nabla\theta = 0, \qquad \nabla\cdot\boldsymbol{u} = 0, \tag{4.1}$$

with initial condition $\theta(\boldsymbol{x}, 0) = \theta_0(\boldsymbol{x}) \in L^2(\Omega)$. Note that (4.1) preserves $\|\theta(\cdot, t)\| = \|\theta_0\|$ for all time, so $\|\theta(\cdot, t)\|$ is uniformly bounded in time by $\|\theta_0\|$. We have left out diffusion in (4.1): in the present section we will discuss how we can define mixing without appealing to diffusion, and how the norms (3.1) and (3.3) can be related to this type of mixing. The definition of mixing we will use is the one from ergodic theory.

The property of mixing in the sense of ergodic theory is a little opaque when described mathematically. The rigorous definition is as follows:

> *Let $(X, \mathcal{A}, \mu)$ be a normalised measure space and $S^t : X \to X$ be a measure-preserving flow. S^t is called* mixing *if*
>
> $$\lim_{t\to\infty} \mu(A \cap S^{-t}(B)) = \mu(A)\,\mu(B), \qquad \textit{for all } A, B \in \mathcal{A}.$$

Here X is our domain Ω, $\mathcal{A}$ is a so-called σ-algebra over X, and μ is a measure. The elements of $\mathcal{A}$ are measurable sets, which we can think of as patches or 'blobs' in Ω. The measure μ assigns a positive real number to a set in $\mathcal{A}$. We have been using Lebesgue measure — the $\mathrm{d}\Omega$ that appears in our integrals — which means the measure of a blob is just its volume. The flow S^t for us takes θ_0 to $\theta(\cdot, t)$ subject to (4.1). The flow S^t preserves volume, and hence Lebesgue measure.

Intuitively, the definition of mixing works as follows. As we only deal with reversible systems here, we can replace $S^{-t}(B)$ by $S^t(B)$ in the definition above, since such a system should be mixing both forward and backward in time (this makes things conceptually easier). Think of A as a fixed reference patch, and B as a blob that gets stirred. Since the transformation is volume-preserving, $S^t(B)$ might stretch and filament, but it does not change its volume. However, if it fills the domain 'uniformly,' then the volume of its intersection with A just ends up being proportional to A and B's volume. Crucially, this is true for every set A and B, so everything ends up everywhere. (Note that this is stronger than ergodicity, which only requires sets to visit every point, but not necessarily be 'everywhere at the same time.')

Mathew *et al* [54,55] have introduced a norm, called the *mix-norm*, which captures the property of mixing in the sense of ergodic theory. The mix-norm is somewhat cumbersome to define on the torus, but to give an idea of its flavor we will describe it for the one-dimensional periodic interval $[0, L]$. First, define

$$d(\theta, x, w) \coloneqq \frac{1}{w} \int_{x-w/2}^{x+w/2} \theta(x') \, \mathrm{d}x' \tag{4.2}$$

for all $x,\ w \in [0, L]$. The function $d(\theta, x, w)$ is the mean value of the concentration θ in an interval of width w centred on x. The mix-norm $\Phi(\theta)$ is then obtained by averaging d^2 over x and w:

$$\Phi^2(\theta) \coloneqq \frac{1}{L^2} \int_0^L \int_0^L d^2(\theta, x, w) \, \mathrm{d}x \, \mathrm{d}w. \tag{4.3}$$

In words, the mix-norm averages the concentration over an interval of width w, then averages the square of this over all intervals and all widths. In dimensions greater than one, the definition of the mix-norm involves integrals over balls of varying sizes instead of intervals.

For our purposes here it suffices that the mix-norm (4.3) and its higher-dimensional generalisation are equivalent to the norm (3.1) with $q = -1/2$ [55]. We will also see below that all Sobolev norms with $q < 0$ capture the property of mixing in the same sense as the mix-norm, as shown by Lin *et al* [50]. For that reason, we extend Mathew *et al* 's terminology and often refer to any negative Sobolev norm as a mix-norm, not just for the case $q = -1/2$. The term *multiscale norm* encompasses the Sobolev norms for any value of q — positive, negative, or zero.

The connection between negative Sobolev norms and mixing involves the property of *weak convergence*. We refer the reader to the book by Lasota & Mackey [48] for a

more complete discussion of the relation between weak convergence and mixing. Weak convergence is defined as follows:

> *A time-dependent function* $f(\boldsymbol{x},t)$, $f(\cdot,t) \in L^2(\Omega)$, *is* weakly convergent *to* $f_\infty \in L^2(\Omega)$ *if*
>
> $$\lim_{t\to\infty} \langle f(\cdot,t), g\rangle = \langle f_\infty, g\rangle, \qquad \textit{for all } g \in L^2(\Omega).$$

Note that if $f(\cdot,t)$ converges weakly to f_∞, then $f(\cdot,t) - f_\infty$ converges weakly to zero.

Instead of a time-dependent function, we can also use a discrete sequence $\{c_m\}$, where m is like time. A simple example of a sequence that converges weakly to zero is given by $c_m = \sin(2\pi m\, x/L)$, since

$$\lim_{m\to\infty} \int_0^L \sin(2\pi m\, x/L)\, g(x)\, \mathrm{d}x = 0$$

for all $g \in L^2(\Omega)$, by the Riemann–Lebesgue lemma. The connection to mixing is that under stirring a passive scalar usually develops finer and finer scales, much like the function $\sin(2\pi m\, x/L)$ with increasing m. In practice diffusion smooths out these large gradients and the concentration field tends to zero at every point. However, if we ignore diffusion and retain the small scales we can still detect this mixing process by 'projecting' onto test functions such as g. We now discuss the connection between mixing in the sense of ergodic theory and the norms (3.3).

The following is a slightly more general version of the mix-norm theorem by Mathew *et al* [55] (the case $q = -1/2$ is equivalent to their theorem):

> *A time-dependent function* $f(\boldsymbol{x},t)$, *where* $f(\cdot,t) \in L^2(\Omega)$ *has mean zero and is bounded in the* L^2 *norm uniformly in time, is weakly convergent to zero if and only if*
>
> $$\lim_{t\to\infty} \|f(\cdot,t)\|_{H^q} = 0, \qquad \textit{for any } q < 0.$$

The proof given by Lin *et al* [50] is reproduced in Appendix A.

Another theorem from Mathew *et al* [55] now implies that the dynamics generated by $\boldsymbol{u}(\boldsymbol{x},t)$ are mixing in the sense of ergodic theory if and only if $\lim_{t\to\infty} \|\theta(\cdot,t)\|_{\dot{H}^q} = 0$, for any $q < 0$. This is a direct consequence of the mix-norm theorem above.

Note that the equivalence of the norms $\|\cdot\|_{\dot{H}^q}$, $q < 0$, with mixing in the sense of ergodic theory is only a useful concept for the freely-decaying problem. In the presence of sources and sinks, diffusion plays the essential role of making an asymptotic state possible (in its absence solutions can diverge), so we cannot simply solve (4.1) with a source term on the right and expect to get anything sensible.

We close this section with a rule of thumb to interpret the decay of mix-norms. Consider a function f with Fourier coefficients $\hat{f}_{\boldsymbol{k}}$, where the coefficients vanish when $\boldsymbol{k}$ contains odd wavenumbers. Now define f' by $\hat{f}'_{\boldsymbol{k}} = \hat{f}_{(\boldsymbol{k}/2)}$, that is, f' is the same as f but with all scales divided by two. From definition (3.4), we have

$$\|f'\|_{\dot{H}^q} = \Big(\sum_{\boldsymbol{k}} k^{2q}\, |\hat{f}_{\boldsymbol{k}/2}|^2\Big)^{1/2} = \Big(\sum_{\boldsymbol{k}'} (2k')^{2q}\, |\hat{f}_{\boldsymbol{k}'}|^2\Big)^{1/2} = 2^q\, \|f\|_{\dot{H}^q}\ . \tag{4.4}$$

Thus, a refinement of scales by a factor of two leads to a decrease in $\|\cdot\|_{\dot{H}^q}$ of a factor 2^q ($q < 0$). For $q = -1$, the norm decreases by half when scales are refined by half. For $q = -1/2$, the *mix-variance* $\|\cdot\|^2_{\dot{H}^q}$ decreases by half when scales are refined by half. In both cases, the norms decrease at a rate which reflects the creation of small scales.

5. Optimisation for decaying problem

5.1. Optimal control

Mathew *et al* [54] have used optimal control techniques [9, 43, 45, 73] to find velocity fields that rapidly reduce the norm $H^{-1/2}$ of a concentration field. (See also [10, 23, 24, 40, 51, 69, 76, 91].) The energy of the flow is held fixed (more precisely, its total action over a time interval). They assume that the velocity field can be expressed as a linear combination of steady incompressible velocity fields $\boldsymbol{u}_i(\boldsymbol{x})$ as

$$\boldsymbol{u}(\boldsymbol{x}, t) = \sum_{i=1}^{n} \alpha_i(t)\, \boldsymbol{u}_i(\boldsymbol{x}), \qquad \nabla \cdot \boldsymbol{u}_i = 0. \tag{5.1}$$

The coefficients $\alpha_i(t)$ the controls that are adjusted to achieve the optimisation. It is assumed that the flow (5.1) can be realised in practice for a given set of functions $\alpha_i(t)$. The quantity to be optimised is the Sobolev norm $\|\theta\|_{H^{-1/2}}$ of a concentration field satisfying the advection equation (1.6), for some initial condition $\theta_0(\boldsymbol{x})$.

In this formalism, the time-integrated energy (i.e., the action A) and the advection equation (1.6) itself enter an augmented functional as constraints:

$$\begin{aligned} W[\boldsymbol{\alpha}, \theta, z, \eta] = \|\theta(\cdot, t_f)\|^2_{H^{-1/2}} - z \left(A - \int_0^{t_f} \boldsymbol{\alpha}(t) \cdot \boldsymbol{R} \cdot \boldsymbol{\alpha}(t)\, \mathrm{d}t \right) \\ - \int_0^{t_f} \left\langle \eta(\boldsymbol{x}, t) \left(\frac{\partial \theta(\boldsymbol{x}, t)}{\partial t} + \sum_{i=1}^{n} \alpha_i(t)\, \boldsymbol{u}_i(\boldsymbol{x}) \cdot \nabla \theta(\boldsymbol{x}, t) \right) \right\rangle \mathrm{d}t, \end{aligned} \tag{5.2}$$

where z and $\eta(\boldsymbol{x}, t)$ are Lagrange multipliers, and the matrix $\boldsymbol{R}$ describes the kinetic energy for the individual velocity fields in (5.1),

$$R_{ij} \coloneqq \tfrac{1}{2} \left\langle u_i(\boldsymbol{x})\, u_j(\boldsymbol{x}) \right\rangle. \tag{5.3}$$

Note that the functional (5.2) involves the norm of the concentration field only at the final time t_f. However, the constraints involve the entire history of $\theta(\boldsymbol{x}, t)$.

Taking the functional (Fréchet) derivatives of (5.2) with respect to $\boldsymbol{\alpha}$, θ, z, η and equating to zero leads to a two-point boundary value problem: $\theta(\boldsymbol{x}, t)$ is specified at the initial time $t = 0$, but $\eta(\boldsymbol{x}, t)$ is specified at the final time t_f (see Eq. (3.2) in [54]). Both θ and η (the 'costate field') satisfy advection equations of the form (1.6), so they can be solved by following particles on Lagrangian trajectories (backwards in time for η), following the velocity field given by the current best guess for $\boldsymbol{\alpha}(t)$. This guess can be varied following an iterative procedure, such as the conjugate gradient method.

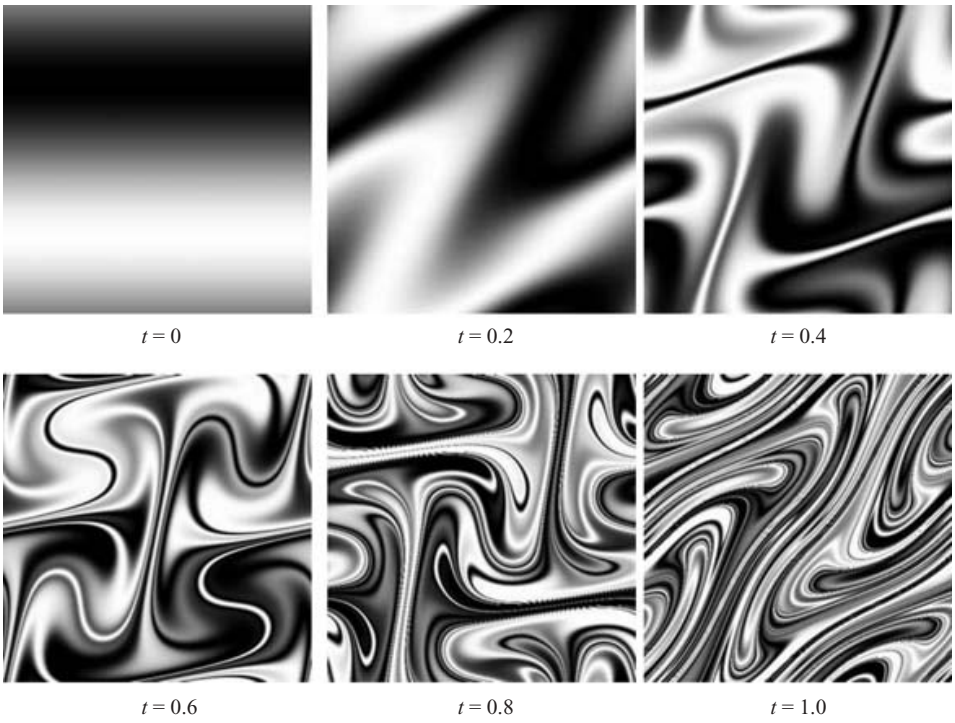


Figure 3. Snapshots of the concentration field $\theta(\boldsymbol{x}, t)$ advected by the optimal solution for the velocity fields defined by (5.4) (from Mathew *et al* [54]).

As an illustration of the method, Mathew *et al* [54] apply it to the velocity fields $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ with streamfunctions

$$\psi_1(x, y) = \sin x \, \sin y, \qquad \psi_2(x, y) = \cos x \, \cos y. \tag{5.4}$$

They set the time interval $t_f = 1$ and action $A = 1/4$, with initial concentration $\theta_0(\boldsymbol{x}) = \sin y$. Figure 3 shows the concentration field evolved by their optimal solution, which is a time-dependent linear combination of (5.4). Note that individually the two flows (5.4) lead to poor mixing, since they are steady and two-dimensional. Figure 3 exhibits finer and finer scales as time evolves, a hallmark of chaotic advection, as well as a roughly exponential decay of the Sobolev norm $\|\theta\|_{H^{-1/2}}$ (Figure 5.2), dotted line).

5.2. Local-in-time optimisation

Lin *et al* [50] proposed an alternative to the global optimal control approach of Mathew *et al* [54]. Instead of focusing on the $\dot{H}^{-1/2}$ norm of the concentration field, they instead examined $\dot{H}^{-1}$. As the theorem in section 4 tells us, any negative Sobolev norm will capture mixing in the sense of ergodic theory, so there is no profound difference in using either norm. However, the rate at which a different norms decrease will in general be different, so different optimal solutions can be obtained. The advantage of $\dot{H}^{-1}$ arises from the evolution equation (3.14), in the absence of sources and diffusion:

$$\frac{d}{dt} \|\theta\|^2_{\dot{H}^{-1}} = 2 \left\langle \nabla^{-1}\theta \cdot \nabla \boldsymbol{u} \cdot \nabla^{-1}\theta \right\rangle . \tag{5.5}$$

The right-hand side is a simple expression that can easily be extremised *instantaneously*, in the sense that given θ at any instant we tweak the velocity to cause the norm to decay as fast as possible. This local-in-time approach can never do better than global optimal control, but it is often good enough, as we will see. But most importantly, it is much less computationally expensive since we do not have to 'peer into the future' and evolve the system forward in time to determine the optimal current velocity field.

In order to formulate an optimisation problem, we must impose some constraints on the velocity field. In section 5.1 we imposed fixed total kinetic energy through (5.3). Now we will consider two types of constraints, fixed energy or fixed power. For a Newtonian fluid, the power is proportional to the L^2 norm $\|\nabla \boldsymbol{u}\|^2$, but we shall refer to this integral as 'the power' even if the fluid is not of this type. The two constraints, then, are to respectively fix

$$\|\boldsymbol{u}\|^2 = U^2 \qquad \text{(fixed energy)} \tag{5.6}$$

or

$$\|\nabla \boldsymbol{u}\|^2 = \sum_{i,j=1}^{d} \left\langle (\partial_i u_j)^2 \right\rangle = \frac{1}{\gamma^2} \qquad \text{(fixed power)}. \tag{5.7}$$

These define the root-mean-square velocity U and rate of strain γ^{-1} of the stirring.

We now proceed with the optimisation technique, that is, to maximise the right-hand side of (5.5). With a few integrations by parts we recast (5.5) in the form

$$\frac{d}{dt} \|\theta\|_{\dot{H}^{-1}}^2 = -2 \left\langle \theta\, \boldsymbol{u} \cdot \nabla(\Delta^{-1}\theta) \right\rangle = -2 \left\langle \boldsymbol{u} \cdot \mathbb{P}(\theta \nabla \phi) \right\rangle \tag{5.8}$$

where ϕ is the filtered scalar field,

$$\phi(\boldsymbol{x}, t) \coloneqq \left(\Delta^{-1}\theta\right)(\boldsymbol{x}, t), \tag{5.9}$$

and $\mathbb{P}(\cdot)$ is the projector onto divergence-free fields defined by

$$\mathbb{P}(\boldsymbol{v}) \coloneqq \boldsymbol{v} - \nabla \Delta^{-1} (\nabla \cdot \boldsymbol{v}). \tag{5.10}$$

Then with either the fixed energy (5.6) or fixed power (5.7) constraint the velocity field maximising the decay rate of $\dot{H}^{-1}$ is

$$\boldsymbol{u}_{\mathrm{e}} = U \frac{\mathbb{P}(\theta \nabla \phi)}{\|\mathbb{P}(\theta \nabla \phi)\|} \qquad \text{(fixed energy)} \tag{5.11}$$

or

$$\boldsymbol{u}_{\mathrm{p}} = -\frac{1}{\gamma} \frac{\Delta^{-1}\mathbb{P}(\theta \nabla \phi)}{\|\mathbb{P}(\theta \nabla \phi)\|_{\dot{H}^{-1}}} \qquad \text{(fixed power)} \tag{5.12}$$

as long as the denominator does not vanish. Hence, $\boldsymbol{u}_{\mathrm{e}}$ or $\boldsymbol{u}_{\mathrm{p}}$ is the best stirring velocity fields to use at any instant in time, unless the denominator vanishes. However, if either

of the norms in the denominators vanishes then $\mathbb{P}(\theta\nabla\phi) = 0$ throughout the domain and *no* incompressible flow can instantaneously decrease the $\dot{H}^{-1}$ norm. For example, this will happen if the concentration field satisfies $\Delta\theta = F(\theta)$, which includes cases where θ is an eigenfunction of the Laplacian. If this situation arises in the course of the time-evolution of θ, then some other optimisation strategy must be adopted.

The most natural alternative when $\mathbb{P}(\theta\nabla\phi) = 0$ is to carry to optimisation to the order, that is, find the velocity field that minimises

$$\frac{d^2}{d^2t}\,\|\theta\|^2_{\dot{H}^{-1}} = 2\left\langle \left[\boldsymbol{u}\cdot\nabla\phi\,\nabla\theta\cdot\boldsymbol{u} - (\boldsymbol{u}\cdot\nabla\theta)\Delta^{-1}(\boldsymbol{u}\cdot\nabla\theta)\right]\right\rangle. \tag{5.13}$$

Then the optimal incompressible flow $\boldsymbol{u}$ solves the eigenvalue problem

$$\Lambda\,\boldsymbol{u} = \mathbb{P}\Big((\boldsymbol{u}\cdot\nabla\theta)\nabla\phi + (\boldsymbol{u}\cdot\nabla\phi)\nabla\theta - 2[\Delta^{-1}(\boldsymbol{u}\cdot\nabla\theta)]\nabla\theta\Big) \tag{5.14}$$

for the fixed energy constraint (5.6) or

$$\Lambda\,\boldsymbol{u} = -\Delta^{-1}\mathbb{P}\Big((\boldsymbol{u}\cdot\nabla\theta)\nabla\phi + (\boldsymbol{u}\cdot\nabla\phi)\nabla\theta - 2[\Delta^{-1}(\boldsymbol{u}\cdot\nabla\theta)]\nabla\theta\Big) \tag{5.15}$$

for the fixed power constraint (5.6). In either case we seek the eigenfunction corresponding to the minimum eigenvalue $\Lambda_- < 0$ to use as the stirring field momentarily, until $\mathbb{P}(\theta\nabla\phi) \neq 0$. The eigenvalue problems in (5.14) and (5.15) are generally difficult; see [50] for a discussion. In practice, we may need to solve one of the eigenvalue problems if we choose a 'bad' initial condition (such as an eigenfunction of the Laplacian, which is commonly done), but one it has started the optimisation procedure does not seem to get stuck very often.

This local-in-time optimal stirring strategy is a limiting case of the short-horizon optimisation studied by Cortelezzi *et al* [23] when the horizon becomes infinitesimal, but the locality and simplicity allows a much broader class of flows to be used. In order to implement it in practice the full scalar field must be monitored so that the optimal flow field can be computed at each instant.

We reproduce here the tests of this optimal stirring strategy presented in Lin *et al* [50]. They used initial scalar distribution $\theta_0(\boldsymbol{x}) = \sin x$ in a domain of size $L = 2\pi$ in $d = 2$ spatial dimensions, for the fixed power constraint (5.7) with $\gamma^{-1} = 6.25 \times (2\pi)^2$, equivalent to the amplitude of the bi-component control used by Mathew *et al* [54]. The results for various norms are shown in Figure 5.2. The optimisation was performed for $\dot{H}^{-1}$, but the $\dot{H}^{-1/2}$ norm is also plotted to allow a direct comparison with Mathew *et al* [54]. The local-time-optimisation seems to outperform the global optimal control, but this is because the former has access to all possible incompressible velocity fields. The difference is evident when comparing figure 3 and figure 5, which shows the concentration field. The Lin *et al* [50] solution uses much smaller velocity scales (though always at fixed power, so the flow must slow down). Note also the the optimal flow in figure 5 is suggestively self-similar in time.

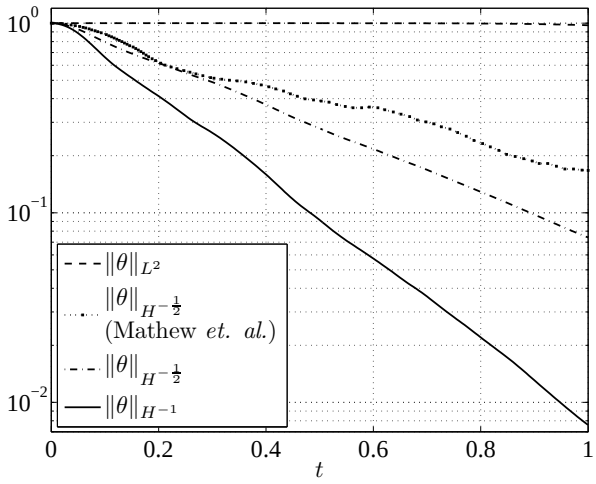


Figure 4. Evolution of norms with the fixed power constraint (5.7) for $\theta_0(\boldsymbol{x}) = \sin x$. All norms are rescaled by their initial values, and the conserved L^2 norm is monitored as a numerical check. The optimisation is over all possible velocity fields satisfying the power constraint, which is why the local-in-time optimisation outperforms the optimal control approach of Mathew *et al* [54]. Snapshots of the velocity field are shown in figure 5, and in figure 3 for the Mathew *et al* solution (from Lin *et al* [50]).

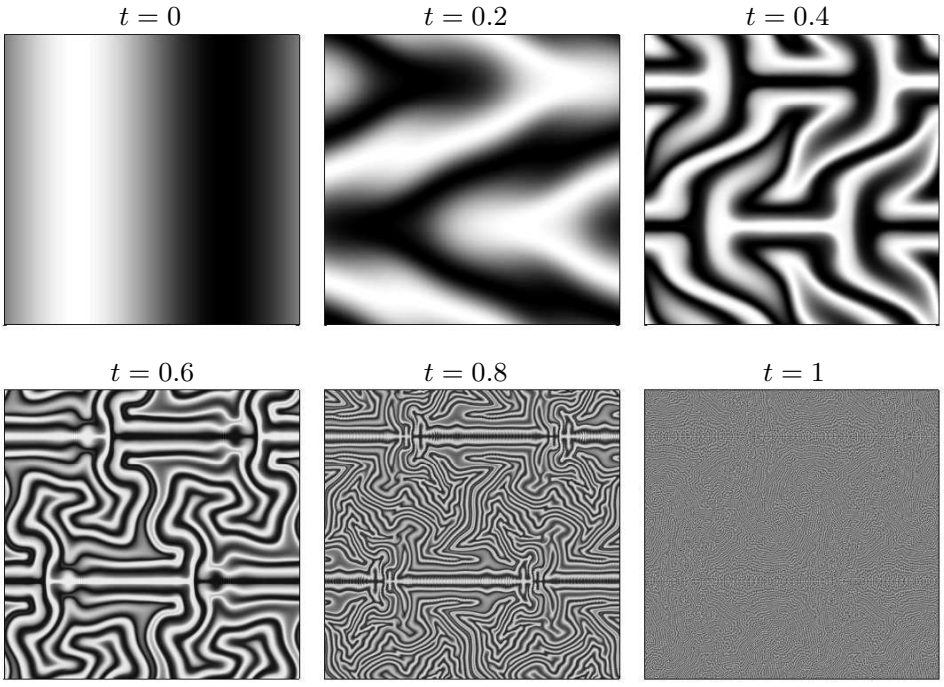


Figure 5. Snapshots of the scalar field in $[0, 2\pi]^2$ for the fixed-power optimal mixing strategy (5.12), with initial condition $\theta_0(\boldsymbol{x}) = \sin x$ (from Lin *et al* [50]).

Part II: The source-sink problem

6. Mixing efficiencies

There are several ways to 'calibrate' the norms to measure some efficiency of mixing. For the freely-decaying case (no sources or sinks), one can normalise the norm by its initial condition to obtain the ratio $\|\theta\|_{\dot{H}^q} / \|\theta_0\|_{\dot{H}^q}$. For the case $q = 0$ this is Danckerts' 'intensity of segregation' [25]. The goal of optimisation is then to reduce $\|\theta\|_{\dot{H}^q} / \|\theta_0\|_{\dot{H}^q}$ as rapidly as possible. This is a strategy that has been used by many authors for the case of variance ($q = 0$) [22] and for negative Sobolev norms [40, 50, 54], as we discussed in section 5.

In the presence of sources and sinks, the norms $\|\theta\|_{\dot{H}^q}$ typically reach an asymptotic steady state (or at least a statistically-steady state). In that case normalising the norms by their initial values is not helpful, since the asymptotic state is usually independent of initial condition. Instead, a convenient measure of mixing efficiency is to normalise the time-asymptotic norm $\|\theta\|_{\dot{H}^q}$ by the value it would have in the absence of stirring. We define *mixing efficiencies* (or *mixing enhancement factors*) by

$$\mathcal{E}_q \coloneqq \|\widetilde{\theta}\|_{\dot{H}^q} / \|\theta\|_{\dot{H}^q} \tag{6.1}$$

where θ is the steady solution to (2.1) and $\widetilde{\theta}$ is the steady solution to the diffusion equation

$$\frac{\partial \widetilde{\theta}}{\partial t} = \kappa \Delta \widetilde{\theta} + s\,. \tag{6.2}$$

(If the velocity field or source are explicitly time-dependent, then an appropriate long-time average must be added to the norms in (6.1); we will see this in section 7.)

The efficiencies measure the amount by which a norm is *decreased* by stirring. If stirring decreases a norm $\|\theta\|_{\dot{H}^q}$ over its purely-diffusive value, then $\mathcal{E}_q$ is larger. For $q \le 0$, an increase in efficiency is associated with better mixing, since the flow has suppressed fluctuations.

We might expect that stirring should always decrease the norms from their purely-diffusive value. For $\mathcal{E}_1$, this is easily shown to be the case [71]. From the definition (3.15) of the inverse gradient of a mean-zero function, we have

$$\begin{aligned} \kappa \|\theta\|_{\dot{H}^1}^2 = \langle \theta s \rangle = \left\langle \theta \nabla \cdot \nabla^{-1} s \right\rangle = - \left\langle \nabla\theta \cdot \nabla^{-1} s \right\rangle \\ \le \|\theta\|_{\dot{H}^1} \left\| \nabla^{-1} s \right\| \end{aligned} \tag{6.3}$$

where we used the Cauchy–Schwarz inequality. The steady-state solution of (6.2) is $\widetilde{\theta} = -\kappa^{-1}\Delta^{-1} s$, so $\nabla\widetilde{\theta} = -\kappa^{-1}\nabla^{-1} s$. We conclude from (6.3) that $\|\theta\|_{\dot{H}^1} \le \|\widetilde{\theta}\|_{\dot{H}^1}$, or

$$\mathcal{E}_1 \ge 1. \tag{6.4}$$

Thus, the efficiency defined with gradients of θ is always decreased by stirring. This is somewhat counter-intuitive, since we expect stirring to create gradients, but this result

holds only for the steady state (or a long-time average). In fact the conventional wisdom in mixing holds that stirring creates sharp gradients, and that those sharp gradients are responsible for good mixing (see introduction). The bound (6.4) shows that this viewpoint must be qualified when sources and sinks are present: stirring may indeed create small scales, but the gradients are never as sharp overall as those that would build up if we didn't stir at all.

Perhaps even more surprising is that the efficiencies $\mathcal{E}_0$ and $\mathcal{E}_{-1}$ are *not* always increased by stirring. The possibility of this was mentioned by Shaw *et al* [71]. Indeed, the combination of flow and source-sink distribution given by§

$$\boldsymbol{u} = (\sin 2x \cos 2y\,,\, -\cos 2x \sin 2y), \tag{6.5a}$$

$$s = \cos 2x \sin y, \tag{6.5b}$$

with $\mathrm{Pe} = 10$ has $\mathcal{E}_0 \simeq .991$, $\mathcal{E}_{-1} \simeq .638$, both less than unity. Here the Péclet number Pe is defined in terms of the L^2 norm of $\boldsymbol{u}$ as

$$\mathrm{Pe} \coloneqq UL/\kappa, \qquad U = \|\boldsymbol{u}\|\,, \tag{6.6}$$

and $L = 2\pi$, $U = 1/\sqrt{2}$ for the velocity field (6.5a). (These unmixing flows may be related to flows that create 'hotspots' [36, 42].)

We can optimise the stirring velocity field to give the *worst* possible mixing efficiency for the source (6.5b), using the same techniques as in section 10.2. Figure 6 shows the resulting streamfunction for $\mathrm{Pe} = 10$, as well as the solutions θ and $\widetilde{\theta}$ to the advection–diffusion and diffusion equation, respectively. The optimised unmixing flow has $\mathcal{E}_0 \simeq .945$, $\mathcal{E}_{-1} \simeq .642$, a modest improvement over (6.5a) (in fact $\mathcal{E}_{-1}$ went up, since the flow was optimised for smallest $\mathcal{E}_0$). At larger Péclet number the optimised solution has lower suboptimal efficiency. The flow appears to achieve this low efficiency by stretching parts of the source along the vertical direction, creating thin peaks, but avoiding concentrating it in other places. Such 'unmixing flows' are fairly rare and delicate to construct. It remains true that for a given source most velocity fields will have efficiency greater than one, though the specific details of this question have not been thoroughly investigated.

In fact it was thought that sources of the form (6.5b), where the source is an eigenfunction of the Laplacian operator, could not lead to $\mathcal{E}_0$ or $\mathcal{E}_{-1}$ less than unity. However, the argument in [71] contains a flaw. To obtain a lower bound on $\mathcal{E}_0$, the authors solved the constrained optimisation problem

$$\langle\theta^2\rangle \le \max_{\vartheta} \{\langle\vartheta^2\rangle \,|\, \kappa\langle|\nabla\vartheta|^2\rangle = \langle s\vartheta\rangle\} \tag{6.7}$$

using the Euler–Lagrange equation

$$2\vartheta_* + 2\mu\kappa\Delta\vartheta_* + \mu s = 0 \tag{6.8}$$

§ The question of whether such 'unmixing' flows exist was posed by Charles R. Doering at the *Workshop on Transport and Mixing in Complex and Turbulent Flows*, Institute for Mathematics and its Applications, Minneapolis, in April 2010. The form (6.5) is derived from a solution suggested by Jeffrey B. Weiss by the end of the workshop.

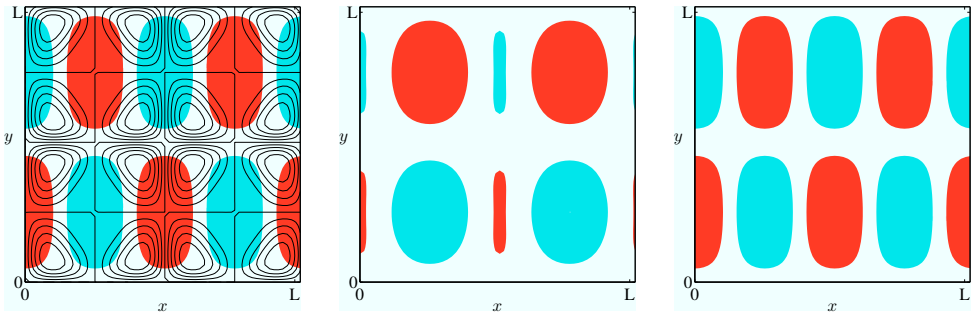


Figure 6. (a) Streamlines of the optimised *unmixing* flow for the source-sink distribution (6.5b) (shown in background) at Pe = 10. (b)–(c) Corresponding temperature field with and without stirring, respectively.

where μ is the Lagrange multiplier enforcing the constraint in (6.7). In terms of the Fourier coefficients the solution of (6.8) is straightforward,

$$\hat{\vartheta}_{*\boldsymbol{k}} = \tfrac{1}{2}\mu\,\hat{s}_{\boldsymbol{k}}/(\mu\kappa k^2 - 1), \tag{6.9}$$

but μ is the solution of the generally-difficult problem

$$\sum_{\boldsymbol{k}} \frac{2 - \mu\kappa k^2}{(\mu\kappa k^2 - 1)^2}\,|\hat{s}_{\boldsymbol{k}}|^2 = 0. \tag{6.10}$$

However, a convexity argument shows that (6.9)–(6.10) is only a maximum if

$$\mu\kappa K^2 \geq 1 \tag{6.11}$$

where $K = 2\pi/L$ is the magnitude of the smallest wavenumber.

When the source is an eigenfunction of the Laplacian, with eigenvalue $-k_s^2$, we can solve for ϑ_* and μ explicitly:

$$\mu = 2/(\kappa k_s^2), \qquad \vartheta_* = s/(\kappa k_s^2), \tag{6.12}$$

for which the maximum criterion (6.11) reads

$$k_s^2 \leq 2K^2. \tag{6.13}$$

Thus, only eigenfunction sources with $k_s^2 = K^2$ and $k_s^2 = 2K^2$ are guaranteed to have $\mathcal{E}_0 \geq 1$. The source (6.5b) has $k_s^2 = 1^2 + 2^2 = 5$ (with $K = 1$), so it can lead to $\mathcal{E}_0 < 1$, as we found numerically. The same criterion (6.13) also hold for $\mathcal{E}_{-1}$ to be bounded below by 1.

Shaw *et al* [71] also derive the rigorous lower bounds

$$\mathcal{E}_0^2 \geq \frac{\sum_{\boldsymbol{k}}(k/K)^{-4}|\hat{s}_{\boldsymbol{k}}|^2}{\sum_{\boldsymbol{k}}(k/K)^{-2}|\hat{s}_{\boldsymbol{k}}|^2}, \qquad \mathcal{E}_{-1}^2 \geq \frac{\sum_{\boldsymbol{k}}(k/K)^{-6}|\hat{s}_{\boldsymbol{k}}|^2}{\sum_{\boldsymbol{k}}(k/K)^{-2}|\hat{s}_{\boldsymbol{k}}|^2}. \tag{6.14}$$

However, these are always less than or equal to unity, so in principle they do not rule out the possibility than any source could be rendered inefficient by *some* flow. Charles R. Doering comments (private communication):

> While the lower bounds (6.14) may be less than one they're greater than zero *uniformly* in Pe. That is, "unstirring" or "herding" is not something that can be enhanced in an unlimited manner by stirring (really a somewhat curious situation in my opinion!).

The bounds (6.14) are also very permissive: for the source (6.5b) they read $\mathcal{E}_0 \geq 1/5$ and $\mathcal{E}_{-1} \geq 1/25$, whereas the optimised unmixing flow in figure 6(a) has $\mathcal{E}_0 \simeq .945$, $\mathcal{E}_{-1} \simeq .642$ (though the unmixing flow could be a local minimum, or could decrease the efficiency further at some higher Péclet number).

7. Upper bounds on mixing efficiencies

It is a simple matter to obtain estimates on the various mixing efficiencies. Thiffeault & Doering [82] used an idea of Doering & Foias [27] to obtain a simple bound on the mixing efficiency $\mathcal{E}_0$: multiply (2.1) by an arbitrary smooth, spatially periodic 'comparison function' $\varphi(\boldsymbol{x})$, integrate, then integrate by parts to find

$$\bigl\langle\!\bigl\langle \theta\,(\boldsymbol{u}\cdot\nabla + \kappa\Delta)\varphi \bigr\rangle\!\bigr\rangle = -\bigl\langle\!\bigl\langle \varphi s \bigr\rangle\!\bigr\rangle. \tag{7.1}$$

Here we introduced the double-bracket notation

$$\bigl\langle\!\bigl\langle F \bigr\rangle\!\bigr\rangle \coloneqq \overline{\langle F(\boldsymbol{x},t)\rangle} \tag{7.2}$$

for a space and time average, the latter defined by

$$\overline{F}(\boldsymbol{x}) \coloneqq \lim_{t\to\infty} \frac{1}{t}\int_0^t F(\boldsymbol{x},t')\,\mathrm{d}t'. \tag{7.3}$$

(We always assume that such time-averages exist.) The time derivative term from the advection-diffusion equation (2.1) has vanished from (7.1), since $\overline{\varphi\,\partial_t\theta} = \overline{\partial_t(\varphi\,\theta)} = \varphi(\boldsymbol{x})\lim_{t\to\infty}\theta(\boldsymbol{x},t)/t = 0$, since θ is bounded. Then apply the Cauchy–Schwarz inequality to (7.1), to obtain

$$\bigl\langle\!\bigl\langle \theta^2 \bigr\rangle\!\bigr\rangle \geq \max_{\varphi} \bigl\langle\!\bigl\langle \varphi s \bigr\rangle\!\bigr\rangle^2 / \bigl\langle\!\bigl\langle (\boldsymbol{u}\cdot\nabla\varphi + \kappa\Delta\varphi)^2 \bigr\rangle\!\bigr\rangle \tag{7.4}$$

where the maximisation is over smooth functions $\varphi(\boldsymbol{x})$. At the cost of some sharpness we can take the square root and then use the Minkowski inequality in the denominator,

$$\bigl\langle\!\bigl\langle \theta^2 \bigr\rangle\!\bigr\rangle^{1/2} \geq \max_{\varphi} |\bigl\langle\!\bigl\langle \varphi s \bigr\rangle\!\bigr\rangle| / \Bigl(\bigl\langle\!\bigl\langle (\boldsymbol{u}\cdot\nabla\varphi)^2 \bigr\rangle\!\bigr\rangle^{1/2} + \kappa\,\|\Delta\varphi\| \Bigr). \tag{7.5}$$

We then apply Hölder's inequality and find

$$\bigl\langle\!\bigl\langle \theta^2 \bigr\rangle\!\bigr\rangle^{1/2} \geq \max_{\varphi} |\bigl\langle\!\bigl\langle \varphi s \bigr\rangle\!\bigr\rangle| / \bigl(U\,\|\nabla\varphi\|_{L^\infty} + \kappa\,\|\Delta\varphi\| \bigr) \tag{7.6}$$

where

$$U = \bigl(\overline{\|\boldsymbol{u}\|^2}\bigr)^{1/2} \tag{7.7}$$

is proportional to the time-averaged total kinetic energy. The two bounds (7.4) and (7.6) both have their uses: the former is tailored to a specific velocity field, but the latter is a global bound valid for *any* stirring velocity field with bounded energy.

To illustrate the usefulness of these estimates, we shall use (7.4) to bound the mixing efficiency $\mathcal{E}_0$, now defined to included a space-time average: where the efficiencies are now defined with a time average,

$$\mathcal{E}_q^2 \coloneqq \overline{\|\widetilde{\theta}\|_{\dot{H}^q}^2} \,/\, \overline{\|\theta\|_{\dot{H}^q}^2}\,. \tag{7.8}$$

Of course, this reduces to the earlier definition (6.1) for time-independent functions. We shall prove the following surprising fact mentioned in the introduction:

> *An optimal way to stir a steady one-dimensional source $s(x)$ in a periodic box, given a fixed time-averaged energy, is to use a spatially-uniform constant flow in the x direction.*

See figure 2 for the type of source-sink distribution and flow that we have in mind. Here, by optimal we mean a flow that maximises (7.8) for $q = 0$, or equivalently minimises the time-averaged variance norm $\langle\!\langle \theta^2 \rangle\!\rangle$.

Now for the proof. First take $\boldsymbol{u}(\boldsymbol{x}, t)$ to be an arbitrary divergence-free vector field. Expand the denominator on the right in (7.4):

$$\left\langle\!\!\left\langle (\boldsymbol{u}\cdot\nabla\varphi + \kappa\Delta\varphi)^2 \right\rangle\!\!\right\rangle = \left\langle\!\!\left\langle (\boldsymbol{u}\cdot\nabla\varphi)^2 \right\rangle\!\!\right\rangle + \kappa^2 \left\langle\!\!\left\langle (\Delta\varphi)^2 \right\rangle\!\!\right\rangle + 2\kappa \left\langle\!\!\left\langle (\boldsymbol{u}\cdot\nabla\varphi)\Delta\varphi \right\rangle\!\!\right\rangle. \tag{7.9}$$

Given that the source $s(\boldsymbol{x}) = s(x)$ is a function of x only, choose $\varphi(\boldsymbol{x}) = \varphi(x)$. Then the last term in (7.9) vanishes:

$$\left\langle\!\!\left\langle (\boldsymbol{u}\cdot\nabla\varphi)\Delta\varphi \right\rangle\!\!\right\rangle = \left\langle\!\!\left\langle u\,\varphi'(x)\varphi''(x) \right\rangle\!\!\right\rangle = \tfrac{1}{2}\left\langle\!\!\left\langle u\,(\varphi'^2)' \right\rangle\!\!\right\rangle = \tfrac{1}{2}\left\langle\!\!\left\langle \boldsymbol{u}\cdot\nabla(\varphi'^2) \right\rangle\!\!\right\rangle = 0. \tag{7.10}$$

We also have $\left\langle\!\!\left\langle (\boldsymbol{u}\cdot\nabla\varphi)^2 \right\rangle\!\!\right\rangle \le U^2 \left\langle |\nabla\varphi|^2 \right\rangle$, where U is defined by (7.7). Hence, from (7.4) we have the bound

$$\left\langle\!\!\left\langle \theta^2 \right\rangle\!\!\right\rangle \ge \max_{\varphi} \left\langle \varphi s \right\rangle^2 / \left(U^2 \left\langle |\nabla\varphi|^2 \right\rangle + \kappa^2 \left\langle (\Delta\varphi)^2 \right\rangle \right). \tag{7.11}$$

We can solve the variational problem (7.11) using its Euler–Lagrange equation, in an identical manner to [28, 71], to find

$$\left\langle\!\!\left\langle \theta^2 \right\rangle\!\!\right\rangle \ge \left\langle s \left\{ \kappa^2\Delta^2 - U^2\Delta \right\}^{-1} s \right\rangle. \tag{7.12}$$

However, the right-hand side of (7.12) is exactly the variance of the periodic zero-mean solution to

$$U\theta'(x) = \kappa\,\theta''(x) + s(x), \tag{7.13}$$

that is, the steady advection-diffusion equation for a constant flow. Hence, the constant flow is optimal, in the sense that any other flow with velocity norm U^2 cannot decrease the variance further. (This optimal solution might not be unique.) This is a surprising

fact: it means that any other process, even turbulence, cannot 'stir' the source-sink better. A constant flow also minimises the $\dot{H}^1$ norm at fixed kinetic energy, since the bound for this norm is

$$\overline{\|\theta\|^2_{\dot{H}^1}} \geq \left\langle s \left\{ \kappa^2 \Delta^2 - U^2 \Delta \right\}^{-1} s \right\rangle \tag{7.14}$$

which is also saturated for the periodic zero-mean solution of (7.13). The third norm, $\dot{H}^{-1}$, associated with $\mathcal{E}_{-1}$, is *not* optimised by a constant flow, reflecting the norm's preference for small scales. A simple bound such as (7.12) and (7.14) which does not depend on the details of $\boldsymbol{u}(\boldsymbol{x},t)$ cannot be derived in this case, except for particular classes of flows (see section 8). Nevertheless, a bound can be derived from [28, 71]

$$\left\langle\!\left\langle |\nabla^{-1}\theta|^2 \right\rangle\!\right\rangle \geq \max_{\varphi} \left\langle\!\left\langle \varphi s \right\rangle\!\right\rangle^2 \Big/ \left\langle\!\left\langle |\nabla \boldsymbol{u} \cdot \nabla\varphi + \boldsymbol{u} \cdot \nabla\nabla\varphi + \kappa \Delta \nabla \varphi|^2 \right\rangle\!\right\rangle ,$$

which after using the Minkowski inequality gives for the denominator gives what is likely a terrible bound:

$$\left\langle\!\left\langle |\nabla^{-1}\theta|^2 \right\rangle\!\right\rangle \geq \max_{\varphi} \left\langle\!\left\langle \varphi s \right\rangle\!\right\rangle^2 \Big/ \left(\gamma^{-1} \left\| \nabla\varphi \right\| + U \left\| \Delta\varphi \right\| + \kappa \left\| \nabla \Delta \varphi \right\| \right)^2 , \tag{7.15}$$

where $\gamma^{-2} = \overline{\|\nabla \boldsymbol{u}\|^2}$ is a time-averaged version of (5.7). This could in principle be maximised over φ, but this is much harder than for the other norms. The important fact about (7.15) is that it depends on the gradient norm $\|\nabla \boldsymbol{u}\|$ of the flow, and clearly the bound can be made arbitrarily small by increasing this norm. Thus flows that minimise this norm are likely mixing.

If one further constrains the problem other solutions are possible (see for example the discussion of Plasting & Young [64] below). Nevertheless, it is surprising that the optimal answer in this case could be so different from a 'mixing' flow, that is, one that amplifies gradients of concentration (in the sense of ergodic theory – see section 4). It is an open question whether there exists source-sink configurations for which the flow that maximises $\mathcal{E}_0$ is also mixing.

To check how sharp the bound (7.4) is for a model system, Thiffeault *et al* [82] considered the two-dimensional 'random sine flow' of Pierrehumbert [2, 62]. This flow consists of alternating horizontal and vertical sine shear flows, with phase angles $\zeta_1(t)$ and $\zeta_2(t) \in [0, 2\pi]$ randomly chosen at each time period, τ. In the first half of the period, the velocity field is

$$\boldsymbol{u}^{(1)}(\boldsymbol{x},t) = \sqrt{2}\, U \left(0 \ , \ \sin(2\pi N x/L + \zeta_1(t))\right); \tag{7.16a}$$

and in the second half-period it is

$$\boldsymbol{u}^{(2)}(\boldsymbol{x},t) = \sqrt{2}\, U \left(\sin(2\pi N y/L + \zeta_2(t)) \ , \ 0\right), \tag{7.16b}$$

where N is an integer indicating the scale of the flow. The source function used is $s(\boldsymbol{x}) = \sqrt{2}\sin(2\pi x/L)$, and we set the integer $N = 1$ for now. For the simple choice $\varphi = s$, we have the efficiency bound

$$\mathcal{E}_0 \leq \sqrt{\frac{\mathrm{Pe}^2}{8\pi^2} + 1} \ . \tag{7.17}$$

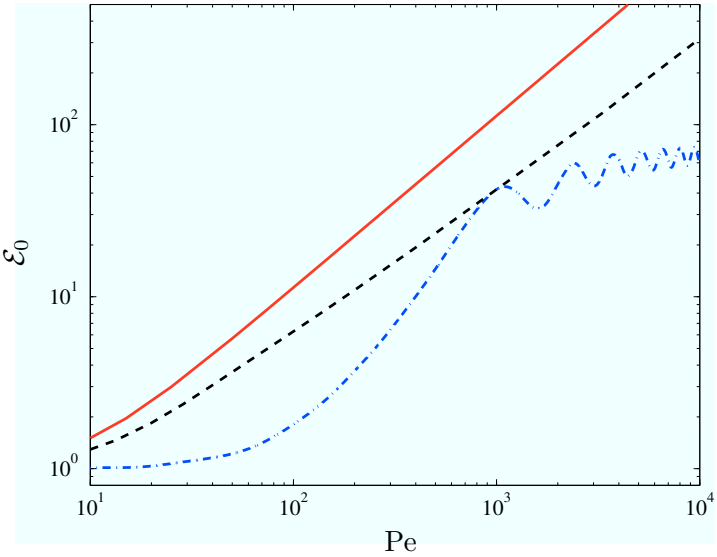


Figure 7. Mixing efficiency $\mathcal{E}_0$ for the random sine flow. The solid line is the upper bound (7.17). The dashed line is the result of direct numerical simulations with U and τ fixed. The dashed-dot curve plots simulation data with κ and τ held constant while varying U (after Thiffeault *et al* [82]).

This bound is plotted in figure 7 against numerical simulations for the random sine flow. The Péclet number is varied in two ways: by varying the diffusivity κ and the amplitude U. The bound captures the trend of the numerical simulation, especially as κ is varied. The oscillations as U gets larger are due to the spatial periodicity of the domain.

Plasting & Young [64] enhanced the bound by including the scalar dissipation rate as a constraint. They define the entropy production (or half the variance dissipation rate) as

$$\chi := \kappa \langle\!\langle |\nabla\theta|^2 \rangle\!\rangle \tag{7.18}$$

which is of course proportional to the time-average of the $\dot{H}^1$-norm of θ. The entropy production satisfies the power integral

$$\chi = \langle\!\langle \theta\, s \rangle\!\rangle. \tag{7.19}$$

Plasting & Young minimise the variance subject to both (7.1) and (7.19), taking χ as given. They find the bound

$$\langle\!\langle \theta^2 \rangle\!\rangle \geq \frac{\langle\!\langle A^2 \rangle\!\rangle \chi^2 + 2\langle\!\langle s\, A \rangle\!\rangle \langle\!\langle s\, \varphi \rangle\!\rangle \chi + \langle\!\langle s\, \varphi \rangle\!\rangle^2 \langle\!\langle s^2 \rangle\!\rangle}{\langle\!\langle A^2 \rangle\!\rangle \langle\!\langle s^2 \rangle\!\rangle - \langle\!\langle s\, A \rangle\!\rangle^2}. \tag{7.20}$$

Their bound takes into account the creation of scalar gradients through the constraint (7.19). For the sine flow (7.16), their bound is plotted in figure 8 in the $\langle\!\langle \theta^2 \rangle\!\rangle$–$\chi$ plane (parabolic solid curve). The horizontal curve at the bottom is the bound (7.17). Notice that, if we know χ, the lower bound of Plasting & Young is a vast improvement over(7.17). The problem is that we usually don't know χ.

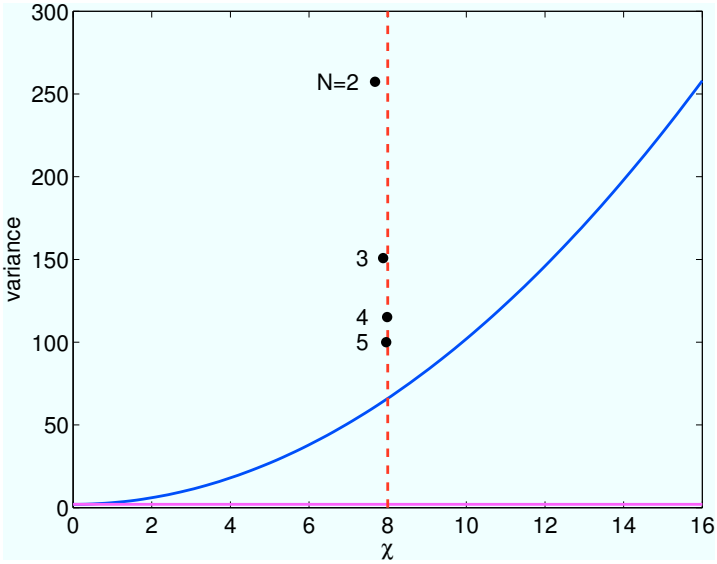


Figure 8. For the sine flow (7.16): the lower bound (7.20) on variance $\langle\!\langle \theta^2 \rangle\!\rangle$ (solid parabolic line) as a function of the dissipation χ. The horizontal solid line is the bound (7.17), which assumes no knowledge of χ. The vertical dashed line is the $N \to \infty$ limit for χ, and the dots are numerical simulation results (after Plasting & Young [64]).

However, for the sine flow we can find χ in the limit $N \to \infty$. First, for the sine flow (7.16) we can compute the effective diffusivity D explicitly [34, 53]:

$$D = \tfrac{1}{8} U^2 \tau \tag{7.21}$$

where we neglect the small molecular diffusivity. For N large in (7.16), we can solve for a 'mean-field' temperature field [64], and obtain

$$\chi \simeq \frac{L^2}{4\pi^2 D} = \frac{2L^2}{\pi^2 U^2 \tau}, \qquad N \gg 1, \tag{7.22}$$

for the same source $s(\boldsymbol{x}) = \sqrt{2}\sin(2\pi x/L)$ used previously. This is the 'homogenisation limit,' where the scale of the source is much larger than the scale of the flow (see section 9). The dashed line in figure 8 shows the large N form (7.22), and the dots are numerical simulations by Plasting & Young for various N ($N = 1$ is off scale). The numerical results approach the vertical line for remarkably small N. They conjecture that for the sine flow as the scale separation (N is made larger) is increased the lower bound is approached. It is an open question whether the bound (7.20) can be realised for more general classes of flow than the single-wavenumber sine flow.

Recently, Alexakis & Tzella [1] addressed the issue of getting bounds that reflect mixing instead of transport in a different way. They focused on a dissipation length scale ℓ_{d} and its inverse k_{d} defined by

$$k_{\mathrm{d}}^2 = \ell_{\mathrm{d}}^{-2} := \overline{\|\theta\|_{\dot{H}^1}^2} / \overline{\|\theta\|^2} = \mathcal{E}_0^2 / \mathcal{E}_1^2 = \chi / (\kappa \langle\!\langle \theta^2 \rangle\!\rangle)\,. \tag{7.23}$$

(This length scale was denoted λ in Thiffeault *et al* [82].) This length scale characterises the scale or variation of the passive scalar. They find that ℓ_{d} scale is *not* always

equivalent to the Batchelor length scale

$$\ell_{\mathrm{B}}^2 := \kappa \ell_u / U \tag{7.24}$$

where ℓ_u is the typical length scale of the velocity field. They introduce the ratio

$$\delta := \ell_u / \ell_s \tag{7.25}$$

where ℓ_s is a typical length scale of the source. For example, for the sine flow (7.16) we have $\ell_u = L/N$, and for the source $s(\boldsymbol{x}) = \sqrt{2}\sin(2\pi x/L)$ we have $\ell_s = L$, so $\delta = N^{-1}$. The homogenisation limit is when $\delta \to 0$, used by Plasting & Young to obtain (7.22). This represents the ideal of scale separation between a small-scale stirring velocity field and a large-scale source. (See section 9.)

Alexakis & Tzella define the correlation $\xi_{\theta,s}$ between the source and the concentration field by

$$\xi_{\theta,s}^2 := \frac{\langle\!\langle \theta\, s \rangle\!\rangle^2}{\langle\!\langle s^2 \rangle\!\rangle \langle\!\langle \theta^2 \rangle\!\rangle}, \qquad 0 \le \xi_{\theta,s} \le 1, \tag{7.26}$$

from which

$$\langle\!\langle \theta^2 \rangle\!\rangle = \xi_{\theta,s}^2\, \ell_{\mathrm{d}}^4\, \kappa^{-2}\, \langle\!\langle s^2 \rangle\!\rangle. \tag{7.27}$$

Given κ and $\langle\!\langle s^2 \rangle\!\rangle$, there are then two ways to reduce the variance: decrease $\xi_{\theta,s}$ or decrease ℓ_{d} (equivalently, increase k_{d}). Decreasing $\xi_{\theta,s}$ is best achieved by *transport*, that is, by having a flow that rapidly carries source onto sink and vice-versa. Decreasing ℓ_{d} relies on *mixing*, that is, by creating small scales of the concentration field. Both achieve the same thing in the end, but in very different ways. Thus, one can target whichever variance-minimising method one prefers by focusing on $\xi_{\theta,s}$ or ℓ_{d}. So far, our emphasis for the advection-diffusion problem with sources and sinks has been on decreasing $\xi_{\theta,s}$. One advantage of aiming instead to decrease ℓ_{d} is that the flows obtained can be good at reducing the variance *regardless of the precise structure of the source-sink configuration.*

A simple bound on k_{d} was given in Thiffeault *et al* [82], and after being adapted to the two scales ℓ_u and ℓ_s it reads

$$k_{\mathrm{d}}^2 \ell_{\mathrm{B}}^2 \le \delta \left(c_1 + c_2 \delta\, \mathrm{Pe}_u^{-1}\right), \qquad \mathrm{Pe}_u := U\ell_u/\kappa, \tag{7.28}$$

where we used a new version of the Péclet number based on the source scale, and the dimensionless constants c_1 and c_2 depend on the shape of the source and velocity field. By working directly from the time-evolution equation for $\nabla\theta$, Alexakis & Tzella [1] improve this to

$$k_{\mathrm{d}}^2 \ell_{\mathrm{B}}^2 \le \tfrac{1}{2} c_3 + \tfrac{1}{2}\sqrt{c_3^2 + 4\delta^3 \mathrm{Pe}^{-1} c_2 (c_1 + c_2 \delta \mathrm{Pe}_u^{-1})}\,. \tag{7.29}$$

where c_3 is a dimensionless constant that depends on the shape of the velocity field. At large δ and large Pe, this is a vast improvement over (7.28), as can be seen in figure 9. The constants were chosen for the sine flow with a sinusoidal source: $c_1 = 2\sqrt{2}$, $c_2 = 2$, $c_3 = \sqrt{2}$ [1]. At smaller δ the crude bound (7.28) does better, which suggests a further

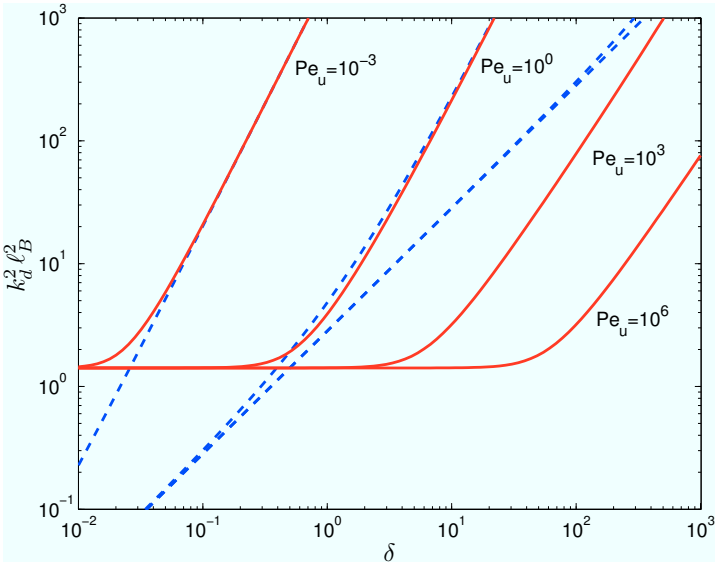


Figure 9. The upper bound (7.28) (dashed line) and the improved bound (7.29) (solid line), for various values of the Péclet number Pe_u. The constants were chosen for the sine flow with a sinusoidal source: $c_1 = 2\sqrt{2}$, $c_2 = 2$, $c_3 = \sqrt{2}$. Note that for $\mathrm{Pe}_u \gtrsim 1$ the bounds (7.28) 'bunch up' and become independent of Pe_u (after Alexakis & Tzella [1]).

improved bound could be derived which captures both. At larger δ the bounds (7.28) 'bunch up' and do not improve further, whereas the bound (7.29) continues to decrease, achieving an asymptotic value $k_\mathrm{d}^2 \ell_\mathrm{B}^2 \leq c_3$ as $\mathrm{Pe}_u \to \infty$.

From the upper bounds (7.28) and (7.29), Alexakis & Tzella [1] investigate the behaviour of k_d as δ varies. They identify five regimes, which are summarised in table 1. In regime I diffusion is very fast, so scalar gradients are set by the source scale and the flow is irrelevant. Regime II is a transitory regime where U and κ now appear explicitly, but the length scale ℓ_u is still absent. Regime III is the classical Batchelor regime [12], where the gradients of θ scale as ℓ_B^{-1}. Regime IV is also Batchelor-like, in that k_d^2 is proportional to U/κ, but where the length scale ℓ_u has been replaced by ℓ_s, since the source now has larger scales than the velocity field. It is indeed remarkable that all these regimes can be captured by (7.28)–(7.29).

There is a fifth regime not captured by these bounds: this is the homogenisation regime when the source has much larger scale than the velocity field ($\delta \ll 1$). We discussed this regime earlier in connection with the Plasting & Young bound; see also Majda & Kramer [53] for an extensive review, or Kramer & Keating [46] and Keating *et al* [44] for a treatment explicitly involving sources and sinks. In homogenisation theory the resulting effective diffusivity D usually scales as $D \sim \mathrm{Pe}^\alpha \kappa$, where $\alpha = 2$ for shear flows (Taylor–Aris dispersion [5, 78]), $\alpha = 1$ for perfect chaotic mixing (D is then independent of κ, as for the sine flow in (7.21)), and $\alpha = 1/2$ for cellular flows. Note that, as pointed out by Alexakis & Tzella, the range of validity in δ of regime V in table 1 may strictly speaking be beyond homogenisation theory: Lin *et al* [49] have

Table 1. The different regimes deduced from the bounds (7.28)–(7.29) (I–IV) and by homogenisation theory (V).

regime	k_{d}^2 estimate	range of validity	note
I	$\leq c_2/\ell_s^2$	$\delta \gg \mathrm{Pe}$	diffusion-dominated
II	$\leq (c_1 c_2 U/\kappa \ell_s^3)^{1/2}$	$\mathrm{Pe}^{1/3} \ll \delta \ll \mathrm{Pe}$	
III	$\leq c_3/\ell_{\mathrm{B}}^2$	$\mathrm{O}(1) \lesssim \delta \ll \mathrm{Pe}^{1/3}$	Batchelor regime
IV	$\leq c_1 U/\kappa \ell_s$	$\delta \lesssim \mathrm{O}(1)$	
V	$\sim \delta^2 \mathrm{Pe}^{\alpha-1}/\ell_{\mathrm{B}}^2$	$\delta \ll \min(1, \mathrm{Pe}^{1-\alpha})$	homogenisation regime

shown that $\delta \ll \mathrm{Pe}^{-1}$ is required for the theory to apply.

8. Dependence of norms on source-sink structure

In this section we discuss the results of Doering & Thiffeault [28] and Shaw, Thiffeault, & Doering [71], who derive bounds on the dependence of the mixing efficiencies with Péclet number. The 'classical' scaling for a smooth source-sink distribution is linear in Pe. But if the source has complicated small-scale structures ('roughness'), then the efficiencies can scale anomalously with Pe, with exponents less than unity, or even logarithmic corrections. The specific behaviour depends on the degree of roughness, as characterised by the rate of decay of the power spectrum for large wavenumbers, as well as the dimensionality of space.

As usual, we consider $\boldsymbol{u}(\boldsymbol{x}, t)$ it to be a specified divergence-free vector field. In addition, we assume the following equal-time single-point statistical properties shared by statistically homogeneous isotropic flows (SHIFs):

$$\begin{aligned} \overline{u_i(\boldsymbol{x},\cdot)} &= 0, \quad & \overline{u_i(\boldsymbol{x},\cdot)u_j(\boldsymbol{x},\cdot)} &= \frac{U^2}{d}\delta_{ij} \\ \overline{u_i(\boldsymbol{x},\cdot)\frac{\partial u_j(\boldsymbol{x},\cdot)}{\partial x_k}} &= 0, \quad & \overline{\frac{\partial u_i(\boldsymbol{x},\cdot)}{\partial x_k}\frac{\partial u_j(\boldsymbol{x},\cdot)}{\partial x_k}} &= \frac{\gamma^2}{d}\,\delta_{ij} \end{aligned} \tag{8.1}$$

where overbar represents the long-time average (assumed to exist) at each point in space. (See [71] for a derivation of these properties.) The r.m.s. velocity U measures the strength of the stirring and γ indicates the flow field's strain or shear content. The ratio $\lambda = U/\gamma$ corresponds to the Taylor microscale for homogeneous isotropic turbulence. The Péclet number for the flow is $\mathrm{Pe} = UL/\kappa$. Note that there are flows, such as the 'random sine flow' (7.16), that satisfy the SHIF conditions (8.1) but are not genuinely isotropic [14, 28, 71]. Nevertheless, we will refer to flows satisfying as SHIFs for expediency.

The reason for introducing SHIFs as a class of flows, in addition to their simplicity and physical relevance, is that the maximisation over φ in (7.4) is particularly simple,

for then the denominator in that equation becomes

$$\left\langle\!\left\langle (\boldsymbol{u}\cdot\nabla\varphi+\kappa\Delta\varphi)^2 \right\rangle\!\right\rangle = \left\langle \frac{U^2}{d}|\nabla\varphi|^2+\kappa^2\left(\Delta\varphi\right)^2 \right\rangle. \tag{8.2}$$

Assuming a time-independent source, the simple variational problem (7.4) then gives [28, 71]

$$\mathcal{E}_0^2 \le \frac{\langle s\,\Delta^{-2}s\rangle}{\left\langle s\left\{\Delta^2-(U^2/\kappa^2 d)\,\Delta\right\}^{-1}s\right\rangle}. \tag{8.3}$$

This bound depends on the spatial structure of the source function, but not its amplitude; the stirring velocity field only enters through the length scale $\kappa/U = \mathrm{Pe}^{-1}L$.

We can bound the small scale and large scale efficiencies $\mathcal{E}_{\pm1}$ from (7.1) in the same manner after integrations by parts and application of the Cauchy–Schwarz inequality. For $\mathcal{E}_1$,

$$\left\langle\!\left\langle \varphi s\right\rangle\!\right\rangle^2 = \left\langle\!\left\langle (\boldsymbol{u}\varphi+\kappa\nabla\varphi)\cdot\nabla\theta\right\rangle\!\right\rangle^2 \le \left\langle\!\left\langle |\boldsymbol{u}\varphi+\kappa\nabla\varphi|^2\right\rangle\!\right\rangle\left\langle\!\left\langle |\nabla\theta|^2\right\rangle\!\right\rangle$$

so

$$\left\langle\!\left\langle |\nabla\theta|^2\right\rangle\!\right\rangle \ge \max_{\varphi}\left\langle\!\left\langle \varphi s\right\rangle\!\right\rangle^2 / \left\langle\!\left\langle (\boldsymbol{u}\varphi+\kappa\nabla\varphi)^2\right\rangle\!\right\rangle. \tag{8.4}$$

A potentially sharper bound involving the full two-point correlation function for the velocity field can be obtained by formally minimising over θ [71], but for our purposes the estimate (8.4) suffices. For SHIFs the denominator in (8.4) is $\langle\varphi[-\kappa^2\Delta+U^2]\varphi\rangle$ and optimisation over φ leads to [28, 71]

$$\mathcal{E}_1^2 \le \frac{\langle s\,(-\Delta^{-1})s\rangle}{\left\langle s\left\{-\Delta+U^2/\kappa^2\right\}^{-1}s\right\rangle} \tag{8.5}$$

for a time-independent source.

We can obtain a bound on $\mathcal{E}_{-1}$ from (7.1) by using $\theta = \nabla\cdot\nabla^{-1}\theta$, integrating by parts, and using Cauchy–Schwarz:

$$\begin{aligned}\left\langle\!\left\langle \varphi s\right\rangle\!\right\rangle^2 &= \left\langle\!\left\langle \nabla(\boldsymbol{u}\cdot\nabla\varphi+\kappa\,\Delta\varphi)\cdot(\nabla^{-1}\theta)\right\rangle\!\right\rangle^2 \\ &\le \left\langle\!\left\langle |\nabla\boldsymbol{u}\cdot\nabla\varphi+\boldsymbol{u}\cdot\nabla\nabla\varphi+\kappa\Delta\nabla\varphi|^2\right\rangle\!\right\rangle\left\langle\!\left\langle |\nabla^{-1}\theta|^2\right\rangle\!\right\rangle\end{aligned}$$

so that

$$\left\langle\!\left\langle |\nabla^{-1}\theta|^2\right\rangle\!\right\rangle \ge \max_{\varphi}\frac{\left\langle\!\left\langle \varphi s\right\rangle\!\right\rangle^2}{\left\langle\!\left\langle |\nabla\boldsymbol{u}\cdot\nabla\varphi+\boldsymbol{u}\cdot\nabla\nabla\varphi+\kappa\Delta\nabla\varphi|^2\right\rangle\!\right\rangle}.$$

For SHIFs the denominator is $\langle\varphi[-\kappa^2\Delta^3+(U^2/d)\,\Delta^2-(\gamma^2/d)\,\Delta]\varphi\rangle$ so that

$$\mathcal{E}_{-1}^2 \le \frac{\langle s\,(-\Delta^{-3})s\rangle}{\left\langle s\left\{-\Delta^3+(U^2/\kappa^2 d)\,\Delta^2-(\gamma^2/\kappa^2 d)\,\Delta\right\}^{-1}s\right\rangle} \tag{8.6}$$

for a time-independent source.

Assuming again that the fluid domain is periodic and that the source is time-independent, it will be helpful to rewrite the mixing efficiency bounds (8.3), (8.5), and (8.6) in Fourier space:

$$\mathcal{E}_1^2 \leq \frac{\sum_{\boldsymbol{k}} |\hat{s}_{\boldsymbol{k}}|^2/k^2}{\sum_{\boldsymbol{k}} |\hat{s}_{\boldsymbol{k}}|^2/(k^2+\mathrm{Pe}^2)}\,, \tag{8.7a}$$

$$\mathcal{E}_0^2 \leq \frac{\sum_{\boldsymbol{k}} |\hat{s}_{\boldsymbol{k}}|^2/k^4}{\sum_{\boldsymbol{k}} |\hat{s}_{\boldsymbol{k}}|^2/(k^4+\mathrm{Pe}^2k^2/d)}\,, \tag{8.7b}$$

$$\mathcal{E}_{-1}^2 \leq \frac{\sum_{\boldsymbol{k}} |\hat{s}_{\boldsymbol{k}}|^2/k^6}{\sum_{\boldsymbol{k}} |\hat{s}_{\boldsymbol{k}}|^2/(k^6+\mathrm{Pe}^2k^4/d+\mathrm{Pe}^2k^2/\lambda^2 d)}\,, \tag{8.7c}$$

where we have rescaled $[0,L]^d$ to $[0,1]^d$ so that wavevector components are integer multiples of 2π. Now we investigate the large Péclet number behaviour of these bounds for a variety of classes of sources.

8.1. Eigenfunction sources.

The simplest class consists of sources that depend only on a single wavenumber k_s, i.e., that are eigenfunctions of the Laplacian Δ with eigenvalue $-k_s^2$. The bounds (8.7) then simplify to

$$\mathcal{E}_1 \leq \sqrt{1+\mathrm{Pe}^2/k_s^2}\,, \tag{8.8a}$$

$$\mathcal{E}_0 \leq \sqrt{1+\mathrm{Pe}^2/k_s^2 d}\,, \tag{8.8b}$$

$$\mathcal{E}_{-1} \leq \sqrt{1+\mathrm{Pe}^2/k_s^2 d+\mathrm{Pe}^2/\lambda^2 k_s^4 d}\,. \tag{8.8c}$$

Observe that each efficiency is asymptotically proportional to Pe for large Pe, corresponding to the expected suppression of variance if the molecular diffusivity κ is replaced by an eddy diffusivity proportional to UL. Moreover these upper bounds are sharp: they may be realised by uniform flow fields whose direction varies appropriately in time to satisfy the weak statistical homogeneity and isotropy conditions used in the analysis [64,71]. Each estimate also exhibits a decreasing dependence on the length scale of the source: for large Pe the bounds for the small- and intermediate-scale efficiencies $\mathcal{E}_1$ and $\mathcal{E}_0$ are proportional to Pe/k_s.

8.2. Square-integrable sources and sinks.

The next simplest case is when the Fourier coefficients of the source-sink distribution are such that the sums in the denominators of (8.7) converge in the limit as $\mathrm{Pe}\to\infty$. For example, the Fourier coefficients of smooth sources decay exponentially for large k, so convergence is guaranteed. We can then use the asymptotic $\mathrm{Pe}\to\infty$ behaviour of

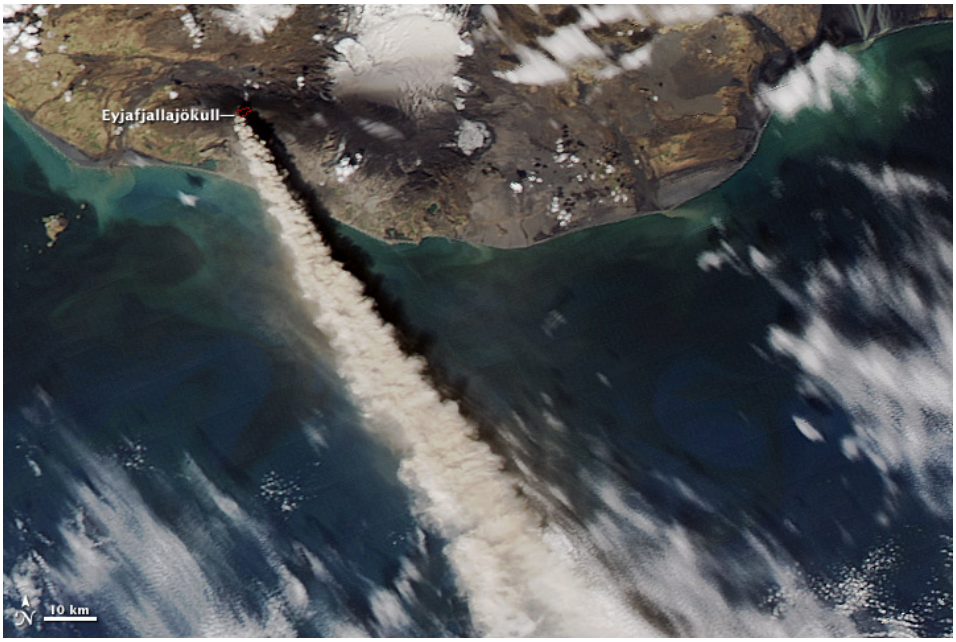


Figure 10. Plume of ash from Eyjafjallajökull volcano, Iceland, 10 May 2010 (NASA MODIS image).

the mixing efficiency bounds to find

$$\mathcal{E}_1 \leq \mathrm{Pe}\sqrt{\frac{\sum_{\boldsymbol{k}}|\hat{s}_{\boldsymbol{k}}|^2/k^2}{\sum_{\boldsymbol{k}}|\hat{s}_{\boldsymbol{k}}|^2}}\,, \tag{8.9a}$$

$$\mathcal{E}_0 \leq \mathrm{Pe}\sqrt{\frac{\sum_{\boldsymbol{k}}|\hat{s}_{\boldsymbol{k}}|^2/k^4}{d\sum_{\boldsymbol{k}}|\hat{s}_{\boldsymbol{k}}|^2/k^2}}\,, \tag{8.9b}$$

$$\mathcal{E}_{-1} \leq \mathrm{Pe}\sqrt{\frac{\sum_{\boldsymbol{k}}|\hat{s}_{\boldsymbol{k}}|^2/k^6}{d\sum_{\boldsymbol{k}}|\hat{s}_{\boldsymbol{k}}|^2/(k^4+k^2/\lambda^2)}}\,. \tag{8.9c}$$

These are the same Pe scalings as above for eigenfunction sources, but the prefactors now depend on different combinations of length scales in the source. For instance, the efficiency $\mathcal{E}_1$ depends more strongly on the high wavenumbers in the source than the other two efficiencies, as would be expected.

8.3. Rough sources.

Rough sources are very common in nature: oil gushing from an underwater well is an infamous recent example of a point source, as is the volcanic ash plume in figure 10. We can define 'rough' sources as those for which some or all the sums in the large-Pe bounds (8.9) are divergent. For instance, if the source is not in L^2 then the denominator in (8.9a) will diverge. For those cases, the Péclet number *scaling* may change, resulting in anomalous behaviour for some or all of the efficiencies (8.7).

The extreme case is the roughest physically-meaningful sources: measure-valued sources such as δ-functions with nondecaying Fourier coefficients $|\hat{s}_{\boldsymbol{k}}| = \mathcal{O}(1)$ as $k \to \infty$.

Then the sums in (8.9a) for $\mathcal{E}_1$ and the denominator of (8.9b) for $\mathcal{E}_0$ diverge in dimension $d = 2$ or 3, rendering those scalings invalid. In this case the Pe dependence of $\mathcal{E}_1$ drops out completely and *all* finite-kinetic-energy stirring velocity fields are completely ineffective at suppressing small scale fluctuations. To determine the high-Pe behaviour of $\mathcal{E}_0$ we approximate sums by integrals. The denominator of (8.7b) is

$$\sum_{\boldsymbol{k}} \frac{1}{k^4 + (\mathrm{Pe}^2/d)\, k^2} \sim \int_{2\pi}^{\infty} \frac{k^{d-1}\,\mathrm{d}k}{k^4 + \mathrm{Pe}^2 k^2/(4\pi^2 d)}. \tag{8.10}$$

For $d = 2$ the integral in (8.10) is

$$\int_{2\pi}^{\infty} \frac{k\,\mathrm{d}k}{k^4 + \mathrm{Pe}^2 k^2/8\pi^2} \sim \frac{\log \mathrm{Pe}}{\mathrm{Pe}^2}\,, \tag{8.11}$$

resulting in the asymptotic bound

$$\mathcal{E}_0 \lesssim \mathrm{Pe}/\sqrt{\log \mathrm{Pe}}\,, \quad d = 2. \tag{8.12}$$

Hence in dimension two there is a logarithmic correction to $\mathcal{E}_0$ as compared to the square-integrable source case.

For $d = 3$ the integral in (8.10) becomes

$$\int_{2\pi}^{\infty} \frac{k^2\,\mathrm{d}k}{k^4 + \mathrm{Pe}^2\, k^2/12\pi^2} \sim \frac{1}{\mathrm{Pe}} \tag{8.13}$$

resulting in an anomalous scaling bound

$$\mathcal{E}_0 \lesssim \sqrt{\mathrm{Pe}}\,, \quad d = 3. \tag{8.14}$$

This is a dramatic modification of the classical scaling. A similar analysis shows that the upper bound on the large scale mixing efficiency $\mathcal{E}_{-1} \sim \mathrm{Pe}$ in (8.9c) persists even for these roughest sources.

We may also analyse anomalous scalings for more general rough sources where the Fourier spectrum $|\hat{s}_{\boldsymbol{k}}|$ decays as $k^{-\beta}$ with $0 \le \beta \le d/2$. The roughest measure–valued sources have $\beta = 0$ while for $\beta > d/2$ the source is square-integrable and thus effectively smooth as far as these multiscale mixing efficiencies are concerned. In order to examine the high-Péclet-number asymptotics of the bounds on the various $\mathcal{E}_q$ we estimate integrals similar to (8.10) but with an extra factor of $k^{-2\beta}$ in the numerator arising from $|\hat{s}_{\boldsymbol{k}}|^2$. The results are summarised in table 2. In $d = 2$ the scaling for $\mathcal{E}_1$ is anomalous for any degree of roughness while $\mathcal{E}_0$ is anomalous only for the roughest sources with $\beta = 0$. In $d = 3$, $\mathcal{E}_1$ is again anomalous for any degree of roughness while $\mathcal{E}_0$ scales anomalously for $0 \le \beta < 1/2$. For both $d = 2$ and 3 the bound on the large scale mixing efficiency $\mathcal{E}_{-1}$ is always classical (i.e., linear in Pe). Of course these scalings neglect any large-k cutoff for the rough sources, as discussed in the next section.

Table 2. Scalings of the bound on the mixing efficiency $\mathcal{E}_q$ as functions of the source roughness exponent β in two and three dimensions.

$d = 2$	$q = 1$	$q = 0$	$q = -1$
$\beta = 0$	1	$\mathrm{Pe}/(\log \mathrm{Pe})^{1/2}$	Pe
$0 < \beta < 1$	Pe^{β}	Pe	Pe
$\beta = 1$	$\mathrm{Pe}/(\log \mathrm{Pe})^{1/2}$	Pe	Pe
$\beta > 1$	Pe	Pe	Pe
$d = 3$			
$\beta = 0$	1	$\mathrm{Pe}^{1/2}$	Pe
$0 < \beta < 1/2$	1	$\mathrm{Pe}^{\beta+1/2}$	Pe
$\beta = 1/2$	1	$\mathrm{Pe}/(\log \mathrm{Pe})^{1/2}$	Pe
$1/2 < \beta < 3/2$	$\mathrm{Pe}^{\beta-1/2}$	Pe	Pe
$\beta = 3/2$	$\mathrm{Pe}/(\log \mathrm{Pe})^{1/2}$	Pe	Pe
$\beta > 3/2$	Pe	Pe	Pe

8.4. Rough sources with a cutoff.

In nature, it can be argued that rough sources are never truly encountered: physical systems tend to be smooth beyond a certain small scale (as long as we stay away from atomic scales, but that is a different story...), or at least they are modelled that way. With this in mind, how are the scalings derived in the previous section realised by sources which are only rough when seen 'from afar' but are actually smooth upon closer examination? Answering this will help, for instance, in understanding how such scalings can be observed in data for which the roughness exponent is meaningful for a limited range of wavenumbers. We will focus on the roughest type of sources for which Fourier coefficients do not decay, but the analysis is easily extended to any type of rough source discussed in section 8.3.

Point-like sources of small but finite size ℓ_s have Fourier coefficients $\hat{s}_{\boldsymbol{k}}$ that are approximately constant in magnitude up to a cutoff wavenumber of order $2\pi/\ell_s$, beyond which the spectrum decays as for a smooth source. We may deduce the behaviour of the bound on $\mathcal{E}_0$ for such sources by inserting an upper limit at $L/\ell_s \gg 1$ into the integral in (8.10). For large but intermediate Péclet numbers satisfying $1 \ll \mathrm{Pe} \ll L/\ell_s$, the cutoff is irrelevant so the logarithmic correction (8.12) in $d = 2$ and the anomalous scaling (8.14) in $d = 3$ appear. However for $\mathrm{Pe} \gg L/\ell_s$, i.e., when the modified Péclet number based on the smallest scale in the source $U\ell_s/\kappa \gg 1$, the smooth source results apply and we recover the mixing efficiency bounds linear in Pe, as in (8.9). Figure (11) shows this scaling transition for the $d = 3$ case. Even in the ultimate regime where the source appears smooth, the *prefactor* in front of the high-Pe scaling bounds are significantly diminished by the small scales in the source: $\mathcal{E}_0 \lesssim [\log(L/\ell_s)]^{-1/2}\,\mathrm{Pe}$ in $d = 2$, and $\mathcal{E}_0 \lesssim [\ell_s/L]^{1/2}\mathrm{Pe}$ in $d = 3$.

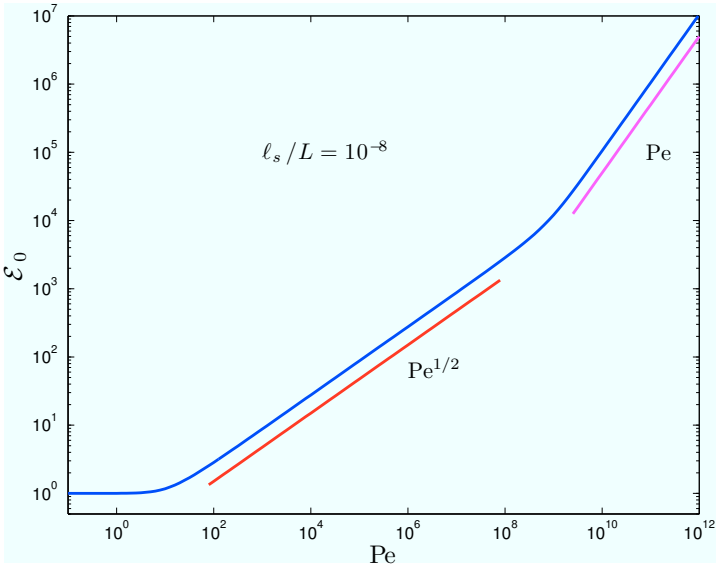


Figure 11. Upper bound for the mixing efficiency $\mathcal{E}_0$ as a function of Péclet number for a small source with $\ell_s = 10^{-8}L$ stirred by a three-dimensional statistically homogeneous and isotropic flow [computed from Eq. (8.7b)]. The intermediate $\mathrm{Pe}^{1/2}$ scaling for $1 \ll \mathrm{Pe} \ll (L/\ell_s)$ is evident (after Doering & Thiffeault [28]).

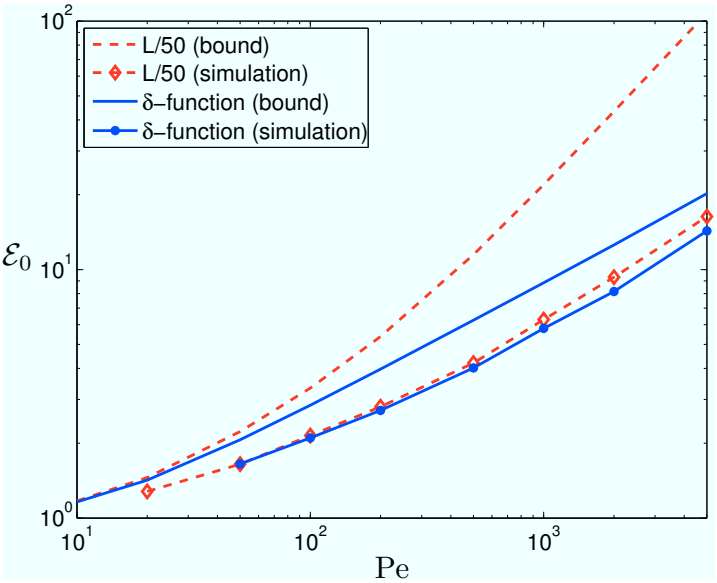


Figure 12. The theoretical upper bounds and simulation results for the mixing enhancement of a cubic source of size $\ell_s = L/50$, and a δ-function source. The stirring velocity field is a 3D version of the random sine flow [62] (after Okabe *et al* [56]).

8.5. Summary and numerical evidence.

An important aspect of the mixing efficiency for rough source-sink distributions discussed above is that an anomalous large-Pe scaling implies that molecular diffusivity is *always* important, for any SHIF. To emphasise: there is no 'residual' effective diffusivity due to stirring in the limit of negligible molecular diffusion. Since we derived upper

bounds, the actual scaling could be worse (see below for numerical results). This again highlights a theme of this review: the source-sink distribution takes centre stage, and should not be treated as a secondary aspect when compared to the stirring flow.

For actual SHIF flows, there is evidence that the upper bound scalings are somewhat generous for finite-size sources. Figure 12 shows results from Okabe *et al* [56], who used a 3D generalisation of Pierrehumbert's random sine flow [62] as a SHIF. The dashed lines are for a source of size $\ell_s = L/50$, where L is the domain size. In that case, the upper bound scales is a rather poor indicator of the actual flow efficiency. However, as the source size is made smaller the bound improves considerably: the solid lines in figure 12 are for a δ-function source. Chertock *et al* [21] have confirmed these results using an accurate operator-splitting method.

Whether the results derived in this section for statistically homogeneous and isotropic flows (SHIFs) generalise to a wider class of flows is an important open question. Another open question is whether there exist SHIFs that saturate the scalings of table 2.

9. Homogenisation theory with sources and sinks

Homogenisation theory is a type of multiscale analysis that exploits a large spatial scale separation between the stirring velocity field and the source [29,34,53,61]. We used this already in describing Plasting & Young's results for the sine flow (7.16) for $N \gg 1$ in section 7, and discussed it in the same section in connection with regime V of Alexakis & Tzella.

Here we review the results of Kramer & Keating [46], Keating *et al* [44], and Lin *et al* [49]. This series of papers has its origin in a definition of an *equivalent diffusivity* defined by Thiffeault *et al* [82] and generalised in Shaw *et al* [71]:

$$D_q^{(\mathrm{eq})} := \kappa\, \mathcal{E}_q \tag{9.1}$$

where $\mathcal{E}_q$ is just the mixing efficiency (7.8). This is the diffusivity that would be required in the absence of stirring to achieve the same level of suppression of the norm $\dot{H}^q$ with stirring. Homogenisation theory has its own *effective diffusivity*, D, which we encountered before, based on a mean-field approach and exploiting separation of scales. (We note that the effective diffusivity is additive on top of the molecular diffusivity, whereas $D_q^{(\mathrm{eq})}$ includes the molecular diffusivity.) For large Péclet number, the effective diffusivity satisfies the rigorous scaling bound [6, 7, 34] $D \le C\,\kappa \mathrm{Pe}^2$, whereas $D_q^{(\mathrm{eq})}$ satisfies $D_q^{(\mathrm{eq})} \le C_2\,\kappa\mathrm{Pe}$ for $q = 1,\ 0,\ -1$ (see table 2). Since both quantities claimed to measure essentially the same thing, both large-Pe scalings could not be right.

The answer, of course, as explained in great detail for a model system by Lin *et al* [49], is that the homogenisation bound applies only when the concept of an effective homogenisation diffusivity exists, that is, in the homogenisation limit $\delta \ll \mathrm{Pe}^{-1}$, where the scale separation $\delta = \ell_u/\ell_s$ was defined in (7.25). Lin *et al* point out that the homogenisation limit $\delta \to 0$ does not commute with the large Pe limit. As a result, the large-Pe dependence of a mixing efficiency such as $\mathcal{E}_0$ has two distinguished

regimes which cross over when δ is of order Pe. They exhibit a specific example where the efficiency $\mathcal{E}_0 \sim \delta^{7/6}\,\mathrm{Pe}^{5/6}$ for fixed δ and $\mathrm{Pe} \to \infty$, consistent with an upper bound linear in Pe. For fixed Pe and $\delta \to 0$, they recover the homogenisation scaling $\mathcal{E}_0 \sim 1 + \mathrm{Pe}^2$. They then introduce a modification of Batchelor's dispersion theory [11], called *Dispersion-Diffusion Theory*, which successfully reconciles effective diffusion in terms of particle dispersion and of suppression of variance of a source-sink distribution.

A general treatment of homogenisation theory with sources and sinks was given by Kramer & Keating [46]. The starting point is a rescaled version of the advection-diffusion equation (2.1): we let $\boldsymbol{x}' = \delta\boldsymbol{x}$, $t' = \delta^2 t$ and immediately drop the primes to get

$$\delta^2\,\frac{\partial\theta}{\partial t} + \delta\,\boldsymbol{u}\left(\frac{\boldsymbol{x}}{\delta}, \frac{t}{\delta^2}\right)\cdot\nabla\theta - \delta^2\,\kappa\Delta\theta = s\left(\frac{\boldsymbol{x}}{\delta}, \frac{t}{\delta^2}\,; \boldsymbol{x}, t\right), \qquad \delta \ll 1, \tag{9.2}$$

where we have assumed that the source can vary on the small spatial scale $\boldsymbol{x}/\delta$ and fast time scale t/δ^2 as well as on the larger spatial scale $\boldsymbol{x}$ and slower time scale t, whereas the velocity field is confined to small spatial scales. The variable $\boldsymbol{x}/\delta$ is assumed periodic, with $\boldsymbol{u}$ having space-time mean zero. The concentration θ is expanded in the usual manner

$$\theta(\boldsymbol{x}, t) = \theta^{(0)}(\boldsymbol{\xi}, \tau; \boldsymbol{x}, t) + \delta\,\theta^{(1)}(\boldsymbol{\xi}, \tau; \boldsymbol{x}, t) + \delta^2\,\theta^{(2)}(\boldsymbol{\xi}, \tau; \boldsymbol{x}, t) + \ldots \tag{9.3}$$

where $\boldsymbol{\xi} := \boldsymbol{x}/\delta$, $\tau := t/\delta^2$ (not to be confused with the period τ of the sine flow (7.16)). With the fast and slow variables separated as in (9.3) we must write $\partial_t \to \delta^{-2}\partial_\tau + \partial_t$, $\nabla \to \delta^{-1}\nabla_{\boldsymbol{\xi}} + \nabla_{\boldsymbol{x}}$. The advection-diffusion operator on the left-hand side of (9.2) then splits into three orders in δ:

$$\mathcal{L}^{(0)} = \partial_\tau + \boldsymbol{u}\cdot\nabla_{\boldsymbol{\xi}} - \kappa\Delta_{\boldsymbol{\xi}}, \qquad \text{at order } \delta^0; \tag{9.4a}$$

$$\mathcal{L}^{(1)} = \boldsymbol{u}\cdot\nabla_{\boldsymbol{x}} - 2\kappa\nabla_{\boldsymbol{x}}\cdot\nabla_{\boldsymbol{\xi}}, \qquad \text{at order } \delta^1; \tag{9.4b}$$

$$\mathcal{L}^{(2)} = \partial_t - \kappa\Delta_{\boldsymbol{x}}, \qquad \text{at order } \delta^2. \tag{9.4c}$$

We have not yet posited a magnitude for the source. The simplest case that will yield a nontrivial self-consistent solution is to take a weak source, $s = \delta^2\hat{s}$, where $\hat{s}$ is order one. Then at leading order we have $\mathcal{L}^{(0)}\theta^{(0)} = 0$, which means that $\theta^{(0)}(\boldsymbol{\xi}, \tau; \boldsymbol{x}, t) = \Theta^{(0)}(\boldsymbol{x}, t)$, i.e., it depends only on the slow variables. (This is a consequence of the solvability condition and uniqueness results — see Lemma 3.1 in [46].) At the next order we have $\mathcal{L}^{(0)}\theta^{(1)} = -\mathcal{L}^{(1)}\theta^{(0)} = -\boldsymbol{u}\cdot\nabla_{\boldsymbol{x}}\Theta^{(0)}$. The solvability condition for this equation says that $\boldsymbol{u}\cdot\nabla_{\boldsymbol{x}}\Theta^{(0)}$ must average to zero over the small scales, which it does since $\boldsymbol{u}$ averages to zero and $\Theta^{(0)}$ does not depend on the small scales. Hence, we have $\theta^{(1)} = \Theta^{(1)}(\boldsymbol{x}, t) + \boldsymbol{\chi}\cdot\nabla_{\boldsymbol{x}}\Theta^{(0)}$, where $\boldsymbol{\chi}$ satisfies the so-called *cell problem*

$$\mathcal{L}^{(0)}\boldsymbol{\chi} = -\boldsymbol{u}(\boldsymbol{\xi}, \tau), \tag{9.5}$$

where $\boldsymbol{\chi}$ has space-time mean zero.

The next and final order has $\mathcal{L}^{(0)}\theta^{(2)} = -\mathcal{L}^{(1)}\theta^{(1)} - \mathcal{L}^{(2)}\theta^{(0)} + \hat{s}$, but we only require the solvability condition that the space-time average of the right-hand side over the small spatial and temporal scales vanishes. This solvability condition leads directly to the homogenised diffusion equation

$$\partial_t \Theta^{(0)} = \nabla_{\boldsymbol{x}} \cdot \big(\mathbb{D}(\boldsymbol{x},t)\,\nabla_{\boldsymbol{x}}\Theta^{(0)}\big) + S(\boldsymbol{x},t) \tag{9.6}$$

where $\mathbb{I}$ is the unit tensor, S is $\hat{s}$ averaged over small space-time scales, and $\mathbb{D}$ is the tensor

$$[\mathbb{D}]_{ij}(\boldsymbol{x},t) \coloneqq \kappa\big(\mathbb{I} + \langle\!\langle \nabla_{\boldsymbol{\xi}}\chi_i \nabla_{\boldsymbol{\xi}}\chi_j \rangle\!\rangle_{\boldsymbol{\xi},\tau}\big), \qquad S(\boldsymbol{x},t) = \langle\!\langle \hat{s} \rangle\!\rangle_{\boldsymbol{\xi},\tau} \tag{9.7}$$

where the subscripts $\boldsymbol{\xi},\tau$ remind us that the average is over small and fast scales $\boldsymbol{\xi}$ and τ. If the system is isotropic, the scalar effective diffusivity D we introduced earlier appears on the diagonal of $\mathbb{D}$.

So far everything has proceeded as one would expect: the homogenised equation (9.6) is exactly the standard one with an averaged source added. How can things go wrong and become more interesting? The most obvious way is if the small-scale average S in (9.7) vanishes identically. In that case we do not get a self-consistent equation involving the source, and we must rescale the source differently. Following Kramer & Keating, we set $s = \delta\hat{s}$, which is a stronger source. At order δ^0 nothing changes from before, but at order δ^1 we get $\mathcal{L}^{(0)}\theta^{(1)} = -\boldsymbol{u}\cdot\nabla_{\boldsymbol{x}}\Theta^{(0)} + \hat{s}$, the source now making an appearance. The solvability condition is still satisfied, since the source averages to zero at the small scales, by assumption. We may then express the solution as

$$\theta^{(1)}(\boldsymbol{\xi},\tau;\boldsymbol{x},t) = \Theta^{(1)}(\boldsymbol{x},t) + \boldsymbol{\chi}\cdot\nabla_{\boldsymbol{x}}\Theta^{(0)} + \theta_s(\boldsymbol{\xi},\tau;\boldsymbol{x},t) \tag{9.8}$$

where $\boldsymbol{\chi}$ is again the unique periodic mean-zero solution to the cell problem (9.5), and $\theta_s(\boldsymbol{\xi},\tau;\boldsymbol{x},t)$ is the unique periodic mean-zero solution to

$$\mathcal{L}^{(0)}\theta_s(\boldsymbol{\xi},\tau;\boldsymbol{x},t) = \hat{s}(\boldsymbol{\xi},\tau;\boldsymbol{x},t) \tag{9.9}$$

for every $(\boldsymbol{x},t)$. Again, this 'source cell problem' has a solution because the source satisfies $\langle\!\langle \hat{s} \rangle\!\rangle_{\boldsymbol{\xi},\tau} = 0$, by assumption. The term θ_s gives us one extra term in the solvability condition for the next order, which now appears as a source instead of the vanishing averaged source in (9.6):

$$\partial_t \Theta^{(0)} = \nabla_{\boldsymbol{x}} \cdot \big(\mathbb{D}(\boldsymbol{x},t)\,\nabla_{\boldsymbol{x}}\Theta^{(0)}\big) - \nabla_{\boldsymbol{x}} \cdot \langle\!\langle \boldsymbol{u}\,\theta_s \rangle\!\rangle_{\boldsymbol{\xi},\tau} \tag{9.10}$$

where $\mathbb{D}$ is defined as in (9.7). Thus in this case because the small-scale mean of the source vanishes the source of concentration arises from the small-scale correlations between $\boldsymbol{u}$ and θ_s that give rise to large-scale variations. To paraphrase Kramer & Keating, θ_s is exactly the local response of the passive scalar field to the source on the small scale, with the large-scale variation frozen at its local value. Then $\boldsymbol{u}\,\theta_s$ is the advective flux of passive scalar density generated in response to the local behaviour of the source. It would be of great interest to generate examples of this type, and to study how their efficiency scales.

The is a third case mentioned by Kramer & Keating [46], when $\nabla_{\boldsymbol{x}} \cdot \langle\!\langle \boldsymbol{u}\, \theta_s \rangle\!\rangle_{\boldsymbol{\xi},\tau}$ vanishes as well as $\langle\!\langle \theta_s \rangle\!\rangle_{\boldsymbol{\xi},\tau}$. Then we must promote the strength of the source again, so that it comes in at order zero. Two source cell problems must then be solved for the, but the resulting homogenised equation looks very similar to (9.10).

The mixing efficiencies associated with (9.7) can easily be derived [44], since in the absence of flow we just set $\mathbb{D} = \kappa$:

$$\mathcal{E}_0^2 = \Big\langle\!\!\Big\langle S \Delta_{\boldsymbol{x}}^{-2} S \Big\rangle\!\!\Big\rangle \Big/ \Big\langle\!\!\Big\langle \left| \left(\nabla_{\boldsymbol{x}} \cdot (\mathbb{D}/\kappa) \cdot \nabla_{\boldsymbol{x}}\right)^{-1} S \right|^2 \Big\rangle\!\!\Big\rangle, \tag{9.11a}$$

$$\mathcal{E}_{-1}^2 = \Big\langle\!\!\Big\langle S \Delta_{\boldsymbol{x}}^{-3} S \Big\rangle\!\!\Big\rangle \Big/ \Big\langle\!\!\Big\langle \left| \nabla^{-1} \left(\nabla_{\boldsymbol{x}} \cdot (\mathbb{D}/\kappa) \cdot \nabla_{\boldsymbol{x}}\right)^{-1} S \right|^2 \Big\rangle\!\!\Big\rangle, \tag{9.11b}$$

where now the space-time averages are over large scales. These simplify considerably if $\mathbb{D}(\boldsymbol{x}, t) = D\,\mathbb{I} = \text{const.}$:

$$\mathcal{E}_0 = D/\kappa, \qquad \mathcal{E}_{-1} = D/\kappa. \tag{9.12}$$

Thus, in this homogenisation limit these two efficiencies are identical, and the definition of equivalent diffusivity (9.1) for $q = -1, 0$ is the same as the effective diffusivity D (if we include the additive molecular value).

The astute reader will have noticed that we did not list $\mathcal{E}_1$. Indeed, Keating *et al* [44] showed that $\mathcal{E}_1$ expressed as for (9.11) directly in terms of $\theta_0 = \Theta_0$ is not correct in the context of homogenisation theory. The cause is that the gradient now takes the scale-separated form $\nabla = \delta^{-1}\nabla_{\boldsymbol{\xi}} + \nabla_{\boldsymbol{x}}$, so the δ^{-1} can promote the smaller-order term θ_1 to leading other. We thus have

$$\nabla\theta = \nabla_{\boldsymbol{x}}\theta^{(0)} + \nabla_{\boldsymbol{\xi}}\theta^{(1)} + \mathrm{O}(\delta^1) \tag{9.13}$$

from which

$$\big\langle\!\!\big\langle |\nabla\theta|^2 \big\rangle\!\!\big\rangle_{\boldsymbol{\xi},\tau} = \big\langle\!\!\big\langle |\nabla_{\boldsymbol{x}}\theta^{(0)}|^2 \big\rangle\!\!\big\rangle_{\boldsymbol{\xi},\tau} + \big\langle\!\!\big\langle |\nabla_{\boldsymbol{\xi}}\theta^{(1)}|^2 \big\rangle\!\!\big\rangle_{\boldsymbol{\xi},\tau} + 2\big\langle\!\!\big\langle \nabla_{\boldsymbol{x}}\theta^{(0)} \cdot \nabla_{\boldsymbol{\xi}}\theta^{(1)} \big\rangle\!\!\big\rangle_{\boldsymbol{\xi},\tau} + \mathrm{O}(\delta^1) \tag{9.14}$$

The last term vanishes since $\nabla_{\boldsymbol{x}}\theta^{(0)}$ doesn't depend on the fast variables, and then using $\theta^{(0)} = \Theta^{(0)}$ and $\theta^{(1)} = \Theta^{(1)}(\boldsymbol{x}, t) + \boldsymbol{\chi} \cdot \nabla_{\boldsymbol{x}}\Theta^{(0)}$ we get

$$\begin{aligned} \big\langle\!\!\big\langle |\nabla\theta|^2 \big\rangle\!\!\big\rangle_{\boldsymbol{\xi},\tau} &= |\nabla_{\boldsymbol{x}}\Theta^{(0)}|^2 + \sum_{i,j} \big\langle\!\!\big\langle \nabla_{\boldsymbol{\xi}}\chi_i \cdot \nabla_{\boldsymbol{\xi}}\chi_j \big\rangle\!\!\big\rangle_{\boldsymbol{\xi},\tau} \frac{\partial\Theta^{(0)}}{\partial x_i} \frac{\partial\Theta^{(0)}}{\partial x_j} \\ &= \nabla_{\boldsymbol{x}}\Theta^{(0)} \cdot (\mathbb{D}/\kappa) \cdot \nabla_{\boldsymbol{x}}\Theta^{(0)} \end{aligned}$$

in the limit as $\delta \to 0$. Inserting the steady solution $\Theta^{(0)} = \left(\nabla_{\boldsymbol{x}} \cdot (\mathbb{D}/\kappa) \cdot \nabla_{\boldsymbol{x}}\right)^{-1} S$, we obtain the gradient norm efficiency

$$\mathcal{E}_1^2 = \Big\langle\!\!\Big\langle S \Delta_{\boldsymbol{x}}^{-1} S \Big\rangle\!\!\Big\rangle \Big/ \Big\langle\!\!\Big\langle \nabla \left(\nabla_{\boldsymbol{x}} \cdot (\mathbb{D}/\kappa) \cdot \nabla_{\boldsymbol{x}}\right)^{-1} S \cdot (\mathbb{D}/\kappa) \cdot \nabla \left(\nabla_{\boldsymbol{x}} \cdot (\mathbb{D}/\kappa) \cdot \nabla_{\boldsymbol{x}}\right)^{-1} S \Big\rangle\!\!\Big\rangle. \tag{9.15}$$

The difference from directly trying to generalise (9.11) for the gradient norm mixing efficiency is that here there is an extra $(\mathbb{D}/\kappa)$ sandwiched in the denominator. It is easier to see how this differs from (9.11) by specialising to $\mathbb{D}(\boldsymbol{x}, t) = D\,\mathbb{I} = \text{const.}$,

$$\mathcal{E}_1 = (D/\kappa)^{1/2}, \tag{9.16}$$

and comparing to (9.12).

Keating *et al* [44] suggest defining a small-scale-averaged version of the equivalent diffusivities,

$$D_q^{(\mathrm{eq})}(\boldsymbol{x},t) \coloneqq \kappa \left\langle\!\!\left\langle |\nabla^q \widetilde{\theta}|^2 \right\rangle\!\!\right\rangle_{\boldsymbol{\xi},\tau} \Big/ \left\langle\!\!\left\langle |\nabla^q \theta|^2 \right\rangle\!\!\right\rangle_{\boldsymbol{\xi},\tau}, \tag{9.17}$$

where as before $\widetilde{\theta}$ is the purely-diffusive solution. The advantage of definition (9.17) over an homogenised D is that it can be made *even if there is no formal scale separation* between the large scales and the small scales, that is, $D_q^{(\mathrm{eq})}(\boldsymbol{x},t)$ characterises scalar dissipation due to processes at 'subgrid scales' whether or not there is a 'gap' between the small and large scales. In fact for $q = \pm 1$ we can go a step further and naturally generalise (9.17) to tensorial quantities,

$$[\mathbb{D}_{\pm 1}^{(\mathrm{eq})}(\boldsymbol{x},t)]_{ij} \coloneqq \kappa \left\langle\!\!\left\langle |\nabla^{\pm 1} \widetilde{\theta}|^2 \right\rangle\!\!\right\rangle_{\boldsymbol{\xi},\tau} \Big/ \left\langle\!\!\left\langle (\nabla^{\pm 1}\theta)_i (\nabla^{\pm 1}\theta)_j \right\rangle\!\!\right\rangle_{\boldsymbol{\xi},\tau}, \tag{9.18}$$

with $q = -1$ being the preferred choice to relate to the effective diffusivity, because of the different scaling for $\mathcal{E}_1$ in (9.16).

10. Optimisation for the source-sink problem

Given sources and sinks, there is an obvious optimisation problem: for fixed energy, which incompressible velocity field has the highest mixing efficiency $\mathcal{E}_q$? The stirring velocity field should in principle satisfy a fluid equation such as Stokes or Navier–Stokes, but we can also optimise over all incompressible velocity fields to get an upper bound on efficiency. This optimisation problem will be discussed in section 10.2.

The presence of sources and sinks implies a different optimisation problem. If the position of the sources and sinks is part of the design process (as it often is for industrial applications), then we may try to optimise the source-sink locations as well. This is a less familiar problem, but one that is easier to tackle because of the structure of (2.1). We discuss thus this problem first in section 10.1.

10.1. Source optimisation

By far the easier optimisation problem is one that is less intuitive: given a stirring velocity field, what is the source-sink distribution which is *best mixed* by the flow. This problem was examined by Thiffeault & Pavliotis [84]. For example, suppose we have a room whose temperature we wish to control, and that there happens to be a predominant airflow in that room which is relatively unaffected by the temperature distribution. Then we can ask where to put heaters (sources) and windows (sinks) so that the temperature is as uniform as possible.

We illustrate the optimisation procedure on the time-independent advection–diffusion equation,

$$\boldsymbol{u}(\boldsymbol{x}) \cdot \nabla\theta - \kappa\Delta\theta = s(\boldsymbol{x}), \qquad \nabla \cdot \boldsymbol{u} = 0, \tag{10.1}$$

in $\Omega = [0,L]^d$ with periodic boundary conditions. The velocity field $\boldsymbol{u}(\boldsymbol{x})$ is specified. Both $\boldsymbol{u}(\boldsymbol{x})$ and $s(\boldsymbol{x})$ are assumed to be sufficiently smooth. As before, we assume that

the source and initial condition have spatial mean zero, which implies that the scalar concentration also has mean zero.

Our goal is to maximise the efficiency $\mathcal{E}_q$ defined by (6.1),

$$\mathcal{E}_q^2 = \|(-\Delta)^{q/2}\widetilde{\theta}\|^2 / \|(-\Delta)^{q/2}\theta\|^2 \,,$$

where $\widetilde{\theta}$ solves equation (10.1) in the absence of advection, $-\kappa\Delta\widetilde{\theta} = s$. In maximising the efficiency $\mathcal{E}_q$, we fix the L^2 norm of the velocity field (or equivalently, the mean kinetic energy of the flow), and vary the diffusivity through the Péclet number $\mathrm{Pe} = \|\boldsymbol{u}\|L/\kappa$.

Define the linear operators

$$\mathcal{L} \coloneqq \boldsymbol{u}(\boldsymbol{x}) \cdot \nabla - \kappa\Delta \qquad \text{and} \qquad \widetilde{\mathcal{L}} \coloneqq -\kappa\Delta,$$

from which we can write the solution to (10.1) and to the purely-diffusive problem as $\theta = \mathcal{L}^{-1}s$ and $\widetilde{\theta} = \widetilde{\mathcal{L}}^{-1}s$, respectively. We can then rewrite the efficiency (6.1) as

$$\mathcal{E}_q^2 = \frac{\|(-\Delta)^{q/2}\widetilde{\mathcal{L}}^{-1}s\|^2}{\|(-\Delta)^{q/2}\mathcal{L}^{-1}s\|^2} = \frac{\left\langle s\,\widetilde{\mathcal{A}}_q^{-1}s\right\rangle}{\left\langle s\,\mathcal{A}_q^{-1}s\right\rangle}\,, \tag{10.2}$$

where the self-adjoint operators $\mathcal{A}_q$ and $\widetilde{\mathcal{A}}_q$ are

$$\mathcal{A}_q \coloneqq \mathcal{L}(-\Delta)^{-q}\mathcal{L}^*\,, \qquad \widetilde{\mathcal{A}}_q \coloneqq \widetilde{\mathcal{L}}(-\Delta)^{-q}\widetilde{\mathcal{L}}^* = \kappa^2(-\Delta)^{2-q}\,, \tag{10.3}$$

and as before $\langle\cdot\rangle$ denotes an average over Ω. To maximise $\mathcal{E}_q^2$, we compute its variation with respect to s and set it equal to zero,

$$\delta\mathcal{E}_q^2 = \frac{2}{\left\langle s\,\mathcal{A}_q^{-1}s\right\rangle}\left\langle\left(\widetilde{\mathcal{A}}_q^{-1}s - \mathcal{E}_q^2\,\mathcal{A}_q^{-1}s\right)\delta s\right\rangle = 0, \tag{10.4}$$

which implies

$$\widetilde{\mathcal{A}}_q^{-1}s = \mathcal{E}_q^2\,\mathcal{A}_q^{-1}s\,. \tag{10.5}$$

This is an eigenvalue problem for the operator $\mathcal{A}_q\widetilde{\mathcal{A}}_q^{-1}$. Optimal sources are given by ground states of the inverse of this operator, and the normalised variance is given by the corresponding (first) eigenvalue. The minimisation problem has a unique minimum, though it may be realised by more than one source-sink distribution, in particular when the flow has symmetry [84].

The operators $\mathcal{A}_q^{-1}$ and $\widetilde{\mathcal{A}}_q^{-1}$ are self-adjoint from $L^2(\Omega)$ to $L^2(\Omega)$; furthermore, they are both positive operators in $L^2(\Omega)$ (restricted to functions with mean zero). Consequently, the generalised eigenvalue problem (10.5) has real positive eigenvalues, and the eigenfunctions s and s' corresponding to distinct eigenvalues are orthogonal with respect to the weighted inner product $(s\,,\,s') \coloneqq \langle s\,\widetilde{\mathcal{A}}_q^{-1}s'\rangle$. For numerical implementation, it is preferable to solve the equivalent self-adjoint eigenvalue problem

$$(\widetilde{\mathcal{A}}_q^{-1/2}\mathcal{A}_q\widetilde{\mathcal{A}}_q^{-1/2})\,r = \mathcal{E}_q^2\,r\,, \qquad s =: \widetilde{\mathcal{A}}_q^{1/2}\,r\,, \tag{10.6}$$

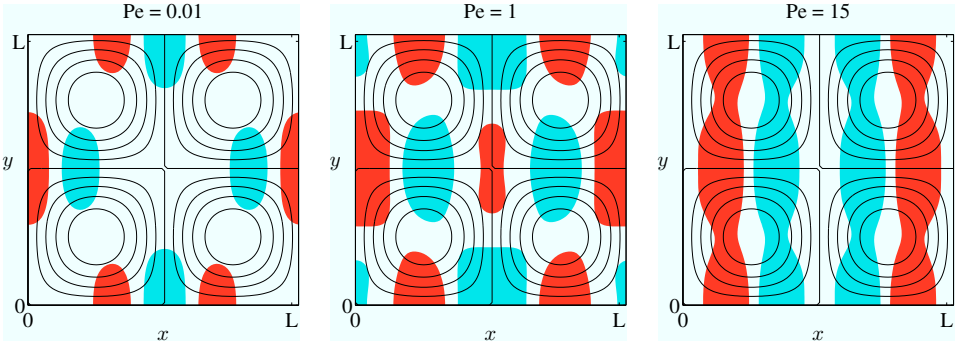


Figure 13. Source distribution that optimises $\mathcal{E}_0$ for the cellular flow (10.7), for three values of the Péclet numbers. Note how there is no source of heat over the stagnation points. The background shading shows hot (red, or dark grey) and cold (blue, or light grey) regions, separated by tepid regions (white). The black contour lines are streamlines of the flow. For both small and large Pe the optimal source converges to an invariant eigenfunction. In all cases there is no source of temperature over the elliptic stagnation points, but in the small Pe case there are sources and sinks over some hyperbolic points.

for the eigenvector r, which then yields the optimal source distribution s. The advantage of the form (10.6) is that the self-adjoint structure of the operator is explicit.

Our goal now is to calculate the optimal source and the corresponding mixing efficiency for some simple velocity fields. Notice that the operator $\widetilde{\mathcal{A}}_q^{-1}$ is a diagonal operator in Fourier space with entries $\kappa^2 k^{2q-4}$, where $k = |\boldsymbol{k}|$ is the magnitude of the wavevector. For $q < 2$, this operator acts as a low-pass filter, suppressing high frequencies.

10.1.1. Dependence on Péclet number We consider the cellular flow on the domain $\Omega = [0, L]^2 = [0, 2\pi]^2$ with streamfunction

$$\psi(x, y) = \sqrt{2} \sin x \sin y \tag{10.7}$$

and velocity field $\boldsymbol{u} = (u_x, u_y) = (\partial_y \psi, -\partial_x \psi)$, normalised to make $\|\boldsymbol{u}\| = 1$. For this cellular flow there are two independent optimal source eigenfunctions with degenerate optimal efficiency; the degeneracy is a consequence of the discrete symmetries of the flow [84]. In figure 13 we show the source that optimises $\mathcal{E}_0$ for three values of Pe. In the foreground are contour lines of the streamfunction. The optimal source distribution appears to become independent of Pe both for small Pe and large Pe, but the distributions are different. The transition between the two regimes occurs when Pe is of order unity. Though the two asymptotic sources are very different, they respect some general principles: the source is arranged for effective transport of hot onto cold and vice versa, and regions of high speed are favoured. In particular, note that the centre of the rolls has a nearly zero, flat source distribution in all cases.

Another perhaps surprising aspect of the small Pe solution in figure 13 is that it has complicated structure. In this large diffusivity limit, one would expect diffusion

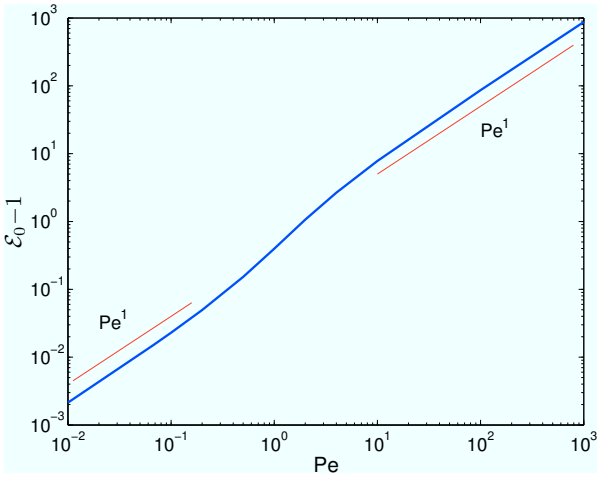


Figure 14. For the flow with streamfunction (10.7), mixing efficiency $\mathcal{E}_0 - 1$ for the optimal source distribution as a function of the Péclet number (after Thiffeault & Pavliotis [84]).

to dominate and gradients to be smoothed out. But since our mixing efficiency (6.1) compares the variance to the unstirred case, which already has very low variance, any amount of improvement will count. Hence, the complicated source for small Pe in figure 13 only gives a minute improvement to the efficiency. The small-Pe optimal solution is particular in that it has some hot and cold spots localised over hyperbolic stagnation points. This is probably due to the high speeds along the separatrices being favoured, even at the cost of straddling hyperbolic stagnation points a little.

In figure 14 we show the value of the optimal efficiency $\mathcal{E}_0$ as a function of Pe. For large Pe, the efficiency typically scales linearly with Pe: this is the 'classical' scaling discussed in [28, 71, 82, 84]. For small Pe, the optimal efficiency also converges towards unity linearly with Pe.

In summary, the optimal source distribution becomes independent of Pe for both large and small Pe, but of course for small Pe the efficiency gain is minimal (since the L^2 norm of the velocity is fixed).

10.1.2. Dependence on norm chosen Our final study will be to examine the behaviour of the optimal efficiency $\mathcal{E}_q$ as q is varied in (6.1). In section 10.1.1 we used $q = 0$; now we fix Pe $= 100$, and allow q to vary over negative and positive values. Figure 15 shows the optimal source distributions for $q = -1$, 0, and 1. For large $|q|$ (not shown), the optimal source distribution converges rapidly to invariant patterns. The $q = -1$ case in figure 15 (negative q) shows small, localised sources and sinks. In contrast, the $q = 1$ case (positive q) shows large, regular localised sources and sinks. In fact, what is striking about the pattern is its simplicity: it is what one might take as a guess at an efficient source distribution, with no added frills. Thus, a high power of q might be useful in situations where a simple configuration is preferable due to engineering constraints. The reason for the simplicity is that spatial variations in the source favour

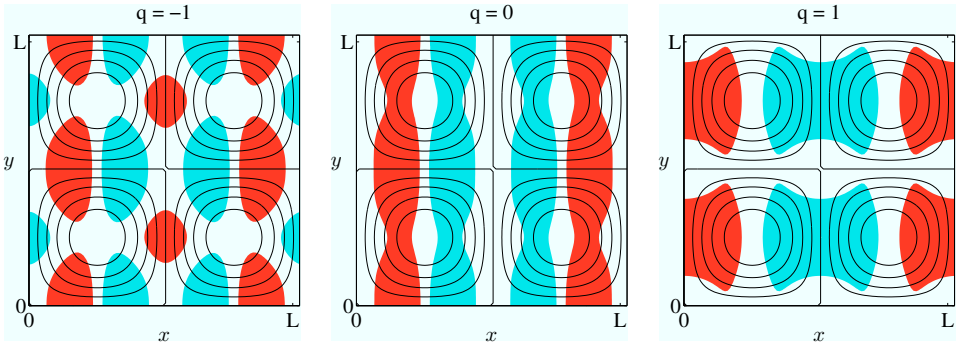


Figure 15. For the flow with streamfunction (10.7), optimal source distribution $s(\boldsymbol{x})$ for Pe $= 100$ and $q = -1$, 0, and 1. In all cases there are no sources or sinks of temperature over the stagnation points. (See the caption to figure 13 for a key to the background shading and contours.)

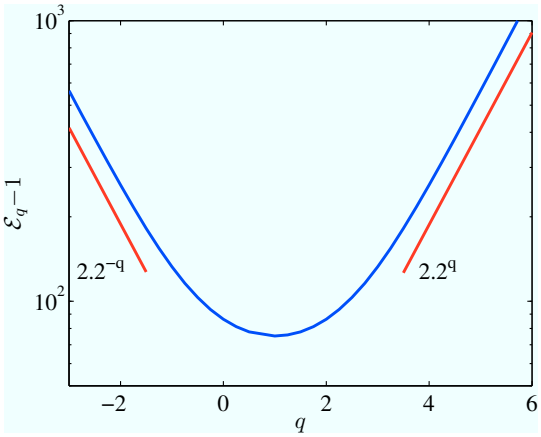


Figure 16. For the flow with streamfunction (10.7), mixing efficiency $\mathcal{E}_q - 1$ as a function of the exponent q for optimal source at Pe $= 100$. The optimal efficiency is symmetric about $q = 1$, and for $|q| \gg 1$ it grows as $(2.2)^{|q|}$ (after Thiffeault & Pavliotis [84]).

the diffusion operator in $\mathcal{L}$, and as $q \to \infty$ these are magnified. Thus, the source must remain as spatially simple as possible while trying to maximise alignment with the velocity. As $q \to -\infty$, spatial variations of the source are downplayed by the norm, allowing more complexity.

Figure 16 shows how the optimal mixing efficiency varies as a function of q. For $|q| \gg 1$, the efficiency scales exponentially as $2.2^{|q|}$. Note that the curve is symmetric about $q = 1$, which leads to a minimum there: whether this is true in general has not been proved, but no counterexample has been found. Thiffeault & Pavliotis [84] provide a partial proof by explicitly finding the symmetry between the operators $\mathcal{A}_{2-q}$ and $\mathcal{A}_q$, but only for small Pe.

10.2. Velocity field optimisation

We turn now to a more obviously relevant problem, that of optimising the stirring velocity field for a given source-sink configuration. As in section 10.1, we restrict to the time-independent problem for simplicity. We will need the functional derivative of $\|\theta\|_{\dot{H}^q}$, which arises from the variation

$$\delta\|\theta\|^2_{\dot{H}^q} = \delta\|(-\Delta)^{q/2}\mathcal{L}^{-1}s\|^2 = \delta\left\langle((-\Delta)^{q/2}\mathcal{L}^{-1}s)^2\right\rangle. \tag{10.8}$$

Since the velocity field only appears in the operator $\mathcal{L}$ defined in (10.1), we have

$$\delta\|\theta\|^2_{\dot{H}^q} = 2\left\langle(-\Delta)^{q/2}\mathcal{L}^{-1}s\,(-\Delta)^{q/2}\delta\mathcal{L}^{-1}s\right\rangle. \tag{10.9}$$

Using the property $\delta\mathcal{L}^{-1} = -\mathcal{L}^{-1}\,\delta\mathcal{L}\,\mathcal{L}^{-1}$ leads to

$$\begin{aligned}\delta\|\theta\|^2_{\dot{H}^q} &= -2\left\langle(-\Delta)^{q/2}\mathcal{L}^{-1}s\,(-\Delta)^{q/2}\mathcal{L}^{-1}\,\delta\mathcal{L}\,\mathcal{L}^{-1}s\right\rangle\\ &= -2\left\langle\left(\mathcal{L}^{-1*}(-\Delta)^q\mathcal{L}^{-1}s\right)\delta\boldsymbol{u}\cdot\nabla\mathcal{L}^{-1}s\right\rangle,\end{aligned}$$

where we integrated by parts, used the adjoint $\mathcal{L}^*$ of $\mathcal{L}$, and substituted $\delta\mathcal{L} = \delta\boldsymbol{u}\cdot\nabla$. From the definition (10.3) of the self-adjoint operator $\mathcal{A}_q$, we can then write the functional derivative as

$$\tfrac{1}{2}\frac{\delta\|\theta\|^2_{\dot{H}^q}}{\delta\boldsymbol{u}} = -\left(\mathcal{A}_q^{-1}s\right)\nabla\mathcal{L}^{-1}s. \tag{10.10}$$

To formulate an optimisation problem, we also need to add constraints on $\boldsymbol{u}$: incompressibility and fixed energy. This is done in the usual manner by considering Lagrange multipliers in the extended functional

$$\mathcal{F}[\boldsymbol{u}] = \tfrac{1}{2}\|\theta\|^2_{\dot{H}^q} + \tfrac{1}{2}\mu\left(\|\boldsymbol{u}\|^2 - U^2\right) + \langle\nu\,\nabla\cdot\boldsymbol{u}\rangle. \tag{10.11}$$

Here μ and $\nu(\boldsymbol{x})$ are the Lagrange multipliers, with ν a function of space since $\nabla\cdot\boldsymbol{u} = 0$ is a pointwise constraint. The functional derivative of $\mathcal{F}[\boldsymbol{u}]$ then gives the Euler–Lagrange equation,

$$\frac{\delta\mathcal{F}[\boldsymbol{u}]}{\delta\boldsymbol{u}} = -\left(\mathcal{A}_q^{-1}s\right)\nabla\mathcal{L}^{-1}s + \mu\,\boldsymbol{u} - \nabla\nu = 0. \tag{10.12}$$

to be solved for $\boldsymbol{u}$ for given s. Note that (10.12) is profoundly nonlinear in $\boldsymbol{u}$, since it enters in the nonlocal operators $\mathcal{L}^{-1}$ and $\mathcal{A}_q^{-1}$.

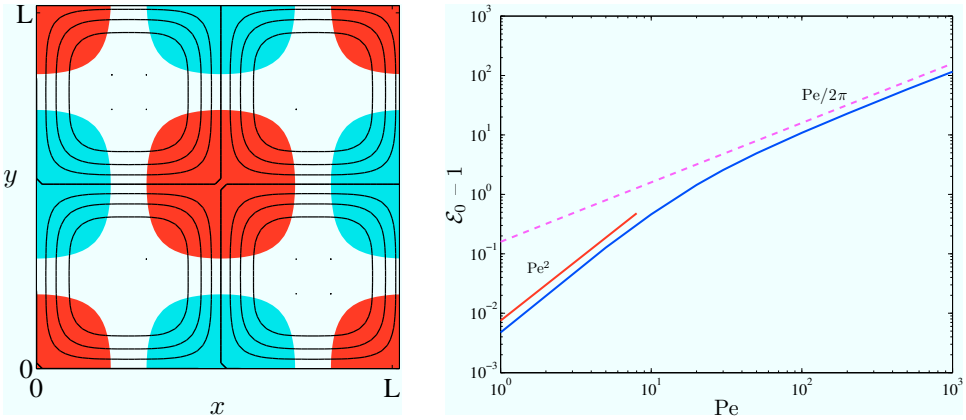
Let us restrict to the two-dimensional case, where we can introduce a streamfunction ψ with $\boldsymbol{u} = (\partial_y\psi, -\partial_x\psi)$. Then taking the curl of (10.12) yields

$$\mu\,\Delta\psi + \left[\mathcal{A}_q^{-1}s\,,\,\mathcal{L}^{-1}s\right] = 0, \tag{10.13}$$

where

$$[f\,,\,g] = \partial_x f\,\partial_y g - \partial_y f\,\partial_x g. \tag{10.14}$$

Equation (10.13) is a nonlinear eigenvalue problem that can be solved in several ways. A direct approach is to start with an initial guess for ψ and μ and compute the *residual*



(a) (b)

Figure 17. (a) Optimal stirring velocity field (streamlines) for the source $\cos x \cos y$, for Pe = 10. (b) Dependence on Péclet number of the optimal mixing efficiency $\mathcal{E}_0$. For small Pe the optimal streamfunction approaches $\sqrt{2}\sin x \sin y$. The dashed line is the upper bound (10.15).

vector, that is the amount by which (10.13) fails to be satisfied. We also append the constraint $(\|\boldsymbol{u}\|^2 - U^2)$ to the residual vector. Then use a multidimensional nonlinear solver such as Matlab's `fsolve` that finds zeroes of the residual vector, by adjusting the vector (ψ, μ). Here ψ has been discretised in some way, either by specifying it on a grid or by expanding it as a Fourier series.

Figure 17(a) shows a solution of (10.13) with $q = 0$, $U = 1$, and Pe = 10 for the source $\cos x \cos y$. The efficiency corresponding to this solution is $\mathcal{E}_0 = 1.46$. The flow is close to the standard cellular flow (10.7), but with a flattened core where velocities are smaller. Whether this is truly optimal is a difficult question: (10.13) has many solutions with different μ, and unlike the source optimisation case (which leads to a linear eigenvalue problem) there is no simple way of finding minimising solutions (but see the upper bound (10.15) below). It is an open challenge to characterise the solutions of (10.12) and (10.13) more thoroughly. The solution in figure 17(a) was obtained from the initial guess $\sqrt{2}\sin x \sin y$, and all other initial conditions examined gave larger values of the norm.

Figure 17(b) shows the dependence on Pe of the optimal mixing efficiency $\mathcal{E}_0$, for the fixed source-sink distribution $\sin x \sin y$ of figure 17(a). For smaller values of Pe, the optimal $\mathcal{E}_0 - 1$ is proportional to Pe^2: this is the diffusion-dominated regime, where stirring only has a small effect. The optimal solution converges to $\sqrt{2}\sin x \sin y$ for small Pe. For larger values of Pe, the optimal solution recovers the 'classical' upper bound scaling, linear in Pe.

It is instructive to compare the optimal efficiency plotted in figure 17(b) with the 'global bound' (7.6). This requires a choice of comparison function φ, and the simplest is to take $\varphi(\boldsymbol{x}) = s(\boldsymbol{x}) = \cos x \cos y$. We then have $\langle \varphi s \rangle = 1/4$, $\|\nabla\varphi\|_{L^\infty} = 1$

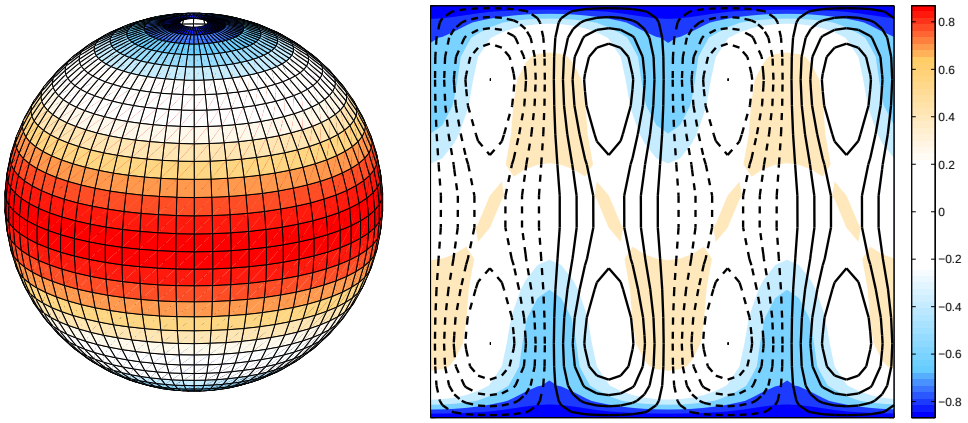


Figure 18. Left: a sphere heated at the equator and cooled at the poles. Right: longitude–latitude plot of the streamlines of a steady flow that maximises $\mathcal{E}_0$, along with the temperature field in the background (from O'Rourke [58]).

and $\|\Delta\varphi\| = 1$ in (7.6). After normalising by the purely-diffusive solution we obtain

$$\mathcal{E}_0 - 1 \le \frac{U}{\kappa} = \frac{\mathrm{Pe}}{2\pi}, \qquad (\text{here } L = 2\pi), \tag{10.15}$$

which is valid even if the velocity field is allowed to be time-dependent, in which case U is defined as in (7.7). The bound (10.15) is plotted as a dashed line in figure 17(b), where we see that our optimal solution is remarkably close to the global upper bound for large Pe (about 37.5% above the optimal solution). This shows that the series of inequalities required to obtain (10.15) do not cause too much loss of sharpness for large Pe, but it also implies that even allowing for arbitrary *time-dependent* stirring cannot improve $\mathcal{E}_0$ by very much. The bound (7.6) thus helps to determine if our local optimal solution is anywhere close to being a global optimum.

11. Discussion

The literature on stirring, mixing, and transport is enormous, and of course we only covered a small corner of it, focusing on direct uses of norms. The decay of variance itself is an object of study, and predicting its decay rate in terms of flow characteristics has been a long-term goal of the theory of mixing. The dominant approaches are the *local theory*, based on dynamical systems quantities such as the statistics of finite-time Lyapunov exponents [2, 3, 8, 33, 72, 79, 88], and the *global theory*, which requires a more thorough analysis of the advection-diffusion equation [35,38,41,52,62,63,65,66,77,80,93]. Neither of these approaches is particularly well-suited to optimisation, so we do not discuss them here. It is also important to note that there are many other measures of mixing beyond norms — see for instance [13, 22, 37, 39, 47, 60, 74, 75, 83, 89, 90] and references therein.

We have reviewed the reasons why various norms were desirable for studying mixing, and how they could be used to optimise the rate of decay for the initial-value, freely-

decaying problem. We then focused for most of the review on the use of norms in the presence of sources and sinks in the long-time limit where the system attain an equilibrium. Our examples used simple periodic geometries.

Let us briefly discuss the work of Constantin *et al* [22], which contains some of the most important recent rigorous results (we leave out a few technical conditions). Their focus is on flows that are *relaxation-enhancing*. These are steady flows with the property that for any given $t > 0$ and $\varepsilon > 0$, it is possible to increase the amplitude of stirring to make $\|\theta(\cdot, t)\| < \varepsilon$. (As usual we assume mean-zero functions.) In other words, it is at least possible to achieve an arbitrary level of mixing, measured according to the L^2 norm, by stirring hard enough. Weakly mixing incompressible flows are always relaxation-enhancing. Constantin *et al* prove that an incompressible velocity field $\boldsymbol{u}(\boldsymbol{x})$ is relaxation-enhancing if and only if $\boldsymbol{u} \cdot \nabla$ has no eigenfunctions in H^1 other than the constant function. Indeed, if there exists such eigenfunctions then most initial conditions will contain some admixture of them, which will not decay. Current examples of relaxation-enhancing flows are, however, not very physical. It is possible that extending the work of Constantin *et al* to time-dependent velocity fields would greatly increase the range of flows that are relaxation-enhancing, but this is likely to be difficult.

There are a number of open problems and areas for further study. We mention a few:

- Find the optimal velocities for more complicated flows, and refine the numerical methods needed to do so; for example O'Rourke has recently examined optimal transport flows on the sphere [58] (see figure 18).
- Understand better the transition between 'transporting' and 'mixing' flows. The flows that minimise the norms in the presence of sources and sinks are very different from mixing flows. Is this a flaw in the measure? What would be a better measure? Are there flows and source-sink distributions which are both optimal in the sense of minimising norms, but are also optimally mixing in the sense of ergodic theory?
- Since all norms $\|\cdot\|_{\dot{H}^q}$ for $q < 0$ act as 'mix-norms,' that is, they decay if the flow is mixing, which one is best? Do we need to select q according to particular applications? Mathew *et al* [54, 55] used $q = -1/2$, and in this review we focused more on $q = -1$, but there is no clear reason to choose either one at this point. On the plus side, however, both work quite well.
- In a similar vein as the previous problem, it is not known whether there are any advantages in using even more general Sobolev norms on the space $W^{q,p}$, $1 \leq p \leq \infty$, rather than on $W^{q,2} = H^q$. Values of p other than 2 are rarely used in the context of mixing, except for $p = \infty$ which is common (see for example [90]). Some rigorous results, such as [22], do not depend on p, which suggests the choice of p matters little.
- Do the results on 'roughness' of the source-sink distribution carry over to open flows, where the sources and sinks can be regarded as distributed on the boundary

rather than in the bulk [81]?

- There has been some work on using the norm approach to quantify mixing in more complex systems with reactions, for example a decaying passive scalar [70], the Fisher equation [14], and the Cahn–Hilliard equation [57]. In this case how do reaction rates, etc., depend on the source-sink structure? What kinds of flows optimise reaction rates?

Acknowledgments

The author is most indebted to Charlie Doering, who co-wrote many papers mentioned in this review. The author also thanks Alexandros Alexakis, Alex Kiselev, Zhi George Lin, George Mathew, Anna Mazzucato, Igor Mezić, Amanda O'Rourke, Tiffany Shaw, Alexandra Tzella, Jeff Weiss, and Andrej Zlatoš for many helpful discussions and comments. This work began while enjoying the hospitality of the Institute for Mathematics and its Applications at the University of Minnesota, and then the Summer Program in Geophysical Fluid Dynamics at the Woods Hole Oceanographic Institution (both supported by NSF). The author was funded by the Division of Mathematical Sciences of the US National Science Foundation, under grant DMS-0806821.

Appendix A. Proof of mix-norm theorem

In this appendix we prove the theorem presented at the end of section 4 relating weak convergence to the negative Sobolev norms. The proof is from Lin *et al* [50]. Write the norm for $H^q(\Omega)$ as

$$\|f\|_{H^q} = \Big(\sum_{\boldsymbol{k}} \lambda_k^{(q)} |\hat{f}_{\boldsymbol{k}}|^2\Big)^{1/2}, \tag{A.1}$$

where $\lambda_k^{(q)} = (1 + k^2 L^2)^q$ for the norm (3.1). Suppose that $f(\cdot, t)$ is uniformly bounded in $L^2(\Omega)$, so that $\|f(\cdot,t)\| \le C$, and $\lim_{t\to\infty} \|f(\cdot,t)\|_{H^q} \to 0$ for some $q < 0$. Then for any $g \in L^2(\Omega)$,

$$\begin{aligned} |\langle f\,,\, g\rangle| &= \left| \sum_{k\le K} \sqrt{\lambda_k^{(q)}}\, \hat{f}_{\boldsymbol{k}}\, \frac{\hat{g}^*_{\boldsymbol{k}}}{\sqrt{\lambda_k^{(q)}}} + \sum_{k>K} \hat{f}_{\boldsymbol{k}}\, \hat{g}^*_{\boldsymbol{k}} \right| \\ &\le \|f\|_{H^q} \left(\sum_{k\le K} \frac{|\hat{g}_{\boldsymbol{k}}|^2}{\lambda_k^{(q)}} \right)^{1/2} + \|f\| \left(\sum_{k>K} |\hat{g}_{\boldsymbol{k}}|^2 \right)^{1/2}. \end{aligned}$$

Given $\epsilon > 0$, first choose $K(\epsilon)$ such that

$$\left(\sum_{k>K(\epsilon)} |\hat{g}_{\boldsymbol{k}}|^2 \right)^{1/2} \le \frac{\epsilon}{2C}, \tag{A.2}$$

then choose $T(\epsilon)$ such that

$$\|f(\cdot,T(\epsilon))\|_{H^q} \le \tfrac{1}{2}\epsilon \left(\sum_{k\le K(\epsilon)} \frac{|\hat{g}_{\boldsymbol{k}}|^2}{\lambda_k^{(q)}} \right)^{-1/2}, \qquad t > T(\epsilon). \tag{A.3}$$

We then have

$$|\langle f\,,\,g\rangle| \le \tfrac{1}{2}\left(1 + C^{-1}\,\|f\|\right)\epsilon \le \epsilon, \qquad t > T(\epsilon),$$

which implies that f converges weakly to zero as $t\to\infty$. (This is true even for $q = 0$.)

Conversely, suppose $\|f(\cdot,t)\| \le C$ for all t and $\lim_{t\to\infty}\langle f\,,\,g\rangle \to 0$ for all $g \in L^2(\Omega)$. By choosing $g = \exp(-i\boldsymbol{k}\cdot\boldsymbol{x})$ we see that all the Fourier coefficients $\hat{f}_{\boldsymbol{k}}(t) \to 0$ as $t\to\infty$. Also, because $\|f(\cdot,t)\|^2 = \sum_{\boldsymbol{k}}|\hat{f}_{\boldsymbol{k}}(t)|^2 \le C^2$ then each $|\hat{f}_{\boldsymbol{k}}(t)| \le C$ for all t.

We have

$$\begin{aligned}\|f\|_{H^q}^2 &= \sum_{k\le K}\lambda_k^{(q)}|\hat{f}_{\boldsymbol{k}}|^2 + \sum_{k>K}\lambda_k^{(q)}|\hat{f}_{\boldsymbol{k}}|^2 \\ &\le \sum_{k\le K}\lambda_k^{(q)}|\hat{f}_{\boldsymbol{k}}|^2 + \lambda_K^{(q)}\,\|f\|^2\,.\end{aligned} \tag{A.4}$$

For any $\epsilon > 0$, we can choose $K(\epsilon)$ such that $\lambda_k^{(q)}\,\|f\| \le \lambda_k^{(q)} C < \epsilon/2$ for $k \ge K(\epsilon)$ (this requires $q < 0$). For any finite K, $\sum_{k\le K}\lambda_k^{(q)}|\hat{f}_{\boldsymbol{k}}(t)|^2 \to 0$ as $t\to\infty$, so there exists $T(\epsilon)$ such that $\sum_{k\le K(\epsilon)}\lambda_k^{(q)}|\hat{f}_{\boldsymbol{k}}(t)|^2 < \epsilon/2$, for all $t > T(\epsilon)$. From (A.4) we obtain $\|f\|_{H^q}^2 < \epsilon$ for all $t > T(\epsilon)$, which proves the result.

References

[1] A. Alexakis and A. Tzella, *Bounding the scalar dissipation scale for mixing flows in the presence of sources*, 2011. arXiv:1103.5011.

[2] T. M. Antonsen, Jr., Z. Fan, E. Ott, and E. Garcia-Lopez, *The role of chaotic orbits in the determination of power spectra*, Phys. Fluids, 8 (1996), pp. 3094–3104.

[3] T. M. Antonsen, Jr. and E. Ott, *Multifractal power spectra of passive scalars convected by chaotic fluid flows*, Phys. Rev. A, 44 (1991), pp. 851–857.

[4] H. Aref, *Stirring by chaotic advection*, J. Fluid Mech., 143 (1984), pp. 1–21.

[5] R. Aris, *On the dispersion of a solute in a fluid flowing through a tube*, Proc. R. Soc. Lond. A, 235 (1956), pp. 66–77.

[6] M. Avellaneda and A. J. Majda, *Stieltjes integral representation and effective diffusivity bounds for turbulent transport*, Phys. Rev. Lett., 62 (1989), pp. 753–755.

[7] ——, *An integral representation and bounds on the effective diffusivity in passive advection by laminar and turbulent flows*, Comm. Math. Phys., 138 (1991), pp. 339–391.

[8] E. Balkovsky and A. Fouxon, *Universal long-time properties of Lagrangian statistics in the Batchelor regime and their application to the passive scalar problem*, Phys. Rev. E, 60 (1999), pp. 4164–4174.

[9] J. Ball, J. E. Marsden, and M. Slemrod, *Controllability for distributed bilinear systems*, SIAM J. Control Optim., 20 (1982), p. 575.

[10] A. Balogh, O. M. Aamo, and M. Krstic, *Optimal mixing enhancement in 3-d pipe flow*, IEEE Trans. Control Sys. Tech., 13 (2005), pp. 27–41.

[11] G. K. Batchelor, *Diffusion in a field of homogeneous turbulence. I. Eulerian analysis*, Australian Journal of Scientific Research, 2 (1949), pp. 437–450.

[12] ——, *Small-scale variation of convected quantities like temperature in turbulent fluid: Part 1. General discussion and the case of small conductivity*, J. Fluid Mech., 5 (1959), pp. 113–133.
[13] D. Bigio and W. Stry, *Measures of mixing in laminar flows*, Polym. Eng. Sci., 30 (1990), pp. 153–161.
[14] D. A. Birch, Y.-K. Tsang, and W. R. Young, *Bounding biomass in the Fisher equation*, Phys. Rev. E, 75 (2007), p. 066304.
[15] M. S. Borgas and B. L. Sawford, *A family of stochastic models for two-particle dispersion in isotropic homogeneous stationary turbulence*, J. Fluid Mech., 279 (1994), pp. 69–99.
[16] M. Chertkov, *Instanton for random advection*, Phys. Rev. E, 55 (1997), pp. 2722–2735.
[17] M. Chertkov, G. Falkovich, and I. Kolokolov, *Intermittent dissipation of a passive scalar in turbulence*, Phys. Rev. Lett., 80 (1998), pp. 2121–2124.
[18] M. Chertkov, G. Falkovich, I. Kolokolov, and V. Lebedev, *Normal and anomalous scaling of the fourth-order correlation function of a randomly advected passive scalar*, Phys. Rev. E, 52 (1995), pp. 4924–4941.
[19] ——, *Statistics of a passive scalar advected by a large-scale two-dimensional velocity field: Analytic solution*, Phys. Rev. E, 51 (1995), pp. 5609–5627.
[20] M. Chertkov, I. Kolokolov, and M. Vergassola, *Inverse cascade and intermittency of passive scalar in one-dimensional smooth flow*, Phys. Rev. E, 56 (1997), pp. 5483–5499.
[21] A. Chertock, C. R. Doering, E. Kashdan, and A. Kurganov, *A fast explicit operator splitting method for passive scalar advection*, J. Sci. Comp., 45 (2010), pp. 200–214.
[22] P. Constantin, A. Kiselev, L. Ryzhik, and A. Zlatoš, *Diffusion and mixing in fluid flow*, Ann. Math., 168 (2008), pp. 643–674.
[23] L. Cortelezzi, A. Adrover, and M. Giona, *Feasibility, efficiency and transportability of short-horizon optimal mixing protocols*, J. Fluid Mech., 597 (2008), pp. 199–231.
[24] D. D'Alessandro, M. Dahleh, and I. Mezić, *Control of mixing in fluid flow: A maximum entropy approach*, IEEE Transactions on Automatic Control, 44 (1999), pp. 1852–1863.
[25] P. V. Danckwerts, *The definition and measurement of some characteristics of mixtures*, Appl. Sci. Res. A, A3 (1952), pp. 279–296.
[26] K. G. Denbigh, *Peter Victor Danckwerts. 14 October 1916–25 October 1984*, Biographical Memoirs of Fellows of the Royal Society, 32 (1986), pp. 98–114.
[27] C. R. Doering and C. Foias, *Energy dissipation in body-forced turbulence*, J. Fluid Mech., 467 (2002), pp. 289–306.
[28] C. R. Doering and J.-L. Thiffeault, *Multiscale mixing efficiencies for steady sources*, Phys. Rev. E, 74 (2006), p. 025301(R).
[29] D. C. P. Donato, *An Introduction to Homogenization*, Oxford University Press, New York, 1999.
[30] I. T. Drummond, *Path-integral methods for turbulent diffusion*, J. Fluid Mech., 123 (1982), pp. 59–68.
[31] P. A. Durbin, *A stochastic model of two-particle dispersion and concentration fluctuations in homogeneous turbulence*, J. Fluid Mech., 100 (1980), pp. 279–302.
[32] C. Eckart, *An analysis of the stirring and mixing processes in incompressible fluids*, J. Mar. Res., 7 (1948), pp. 265–275.
[33] G. Falkovich, K. Gawędzki, and M. Vergassola, *Particles and fields in turbulence*, Rev. Mod. Phys., 73 (2001), pp. 913–975.
[34] A. Fannjiang and G. Papanicolaou, *Convection enhanced diffusion for periodic flows*, SIAM J. Appl. Math., 54 (1994), pp. 333–408.
[35] D. R. Fereday, P. H. Haynes, A. Wonhas, and J. C. Vassilicos, *Scalar variance decay in chaotic advection and Batchelor-regime turbulence*, Phys. Rev. E, 65 (2002), p. 035301(R).
[36] M. D. Finn and S. M. Cox, *Shear flow suppression of hotspot growth*, Physica D, 240 (2011), pp. 709–718.
[37] M. D. Finn, S. M. Cox, and H. M. Byrne, *Mixing measures for a two-dimensional chaotic Stokes flow*, J. Eng. Math., 48 (2004), pp. 129–155.

[38] A. D. Gilbert, *Advected fields in maps. III. Decay of passive scalar in baker's maps*, Dynam. Sys., 21 (2006), pp. 25–71.

[39] E. Gouillart, O. Dauchot, and J.-L. Thiffeault, *Measures of mixing quality in open flows with chaotic advection*, Phys. Fluids, 23 (2010), p. 013604.

[40] O. Gubanov and L. Cortelezzi, *Towards the design of an optimal mixer*, J. Fluid Mech., 651 (2010), pp. 27–53.

[41] P. H. Haynes and J. Vanneste, *What controls the decay of passive scalars in smooth flows?*, Phys. Fluids, 17 (2005), p. 097103.

[42] G. Iyer, A. Novikov, L. Ryzhik, and A. Zlatoš, *Exit times of diffusions with incompressible drift*, SIAM J. Math. Anal., 42 (2010), pp. 2484–2498.

[43] V. Jurdjevic and J. P. Quinn, *Controllability and stability*, J. Diff. Eqns., 28 (1978), pp. 381–389.

[44] S. R. Keating, P. R. Kramer, and K. S. Smith, *Homogenization and mixing measures for a replenishing passive scalar field*, Phys. Fluids, 22 (2010), p. 075105.

[45] D. E. Kirk, *Optimal control theory – An introduction*, Prentice-Hall, Englewood Cliffs, N. J., 1970.

[46] P. R. Kramer and S. R. Keating, *Homogenization theory for a replenishing passive scalar field*, Chin. Ann. Math., Ser. B, 30 (2009), pp. 631–644.

[47] T. S. Krasnopolskaya, V. V. Meleshko, G. W. M. Peters, and H. E. H. Meijer, *Mixing in Stokes flow in an annular wedge cavity*, Eur. J. Mech. B/Fluids, 18 (1999), pp. 793–822.

[48] A. Lasota and M. C. Mackey, *Chaos, Fractals, and Noise*, Springer-Verlag, New York, 1994.

[49] Z. Lin, K. Boďová, and C. R. Doering, *Models & measures of mixing & effective diffusion*, Discr. Cont. Dyn. Sys., 28 (2010), pp. 259–274.

[50] Z. Lin, C. R. Doering, and J.-L. Thiffeault, *An optimal stirring strategy for passive scalar mixing*, 2011. in press.

[51] W. Liu, *Mixing enhancement by optimal flow advection*, SIAM J. Control Optim., 47 (2008), pp. 624–638.

[52] W. Liu and G. Haller, *Strange eigenmodes and decay of variance in the mixing of diffusive tracers*, Physica D, 188 (2004), pp. 1–39.

[53] A. J. Majda and P. R. Kramer, *Simplified models for turbulent diffusion: Theory, numerical modelling and physical phenomena*, Physics Reports, 314 (1999), pp. 237–574.

[54] G. Mathew, I. Mezić, S. Grivopoulos, U. Vaidya, and L. Petzold, *Optimal control of mixing in Stokes fluid flows*, J. Fluid Mech., 580 (2007), pp. 261–281.

[55] G. Mathew, I. Mezić, and L. Petzold, *A multiscale measure for mixing*, Physica D, 211 (2005), pp. 23–46.

[56] T. Okabe, B. Eckhardt, J.-L. Thiffeault, and C. R. Doering, *Mixing effectiveness depends on the source–sink structure: Simulation results*, Journal of Statistical Mechanics: Theory and Experiment, 2008 (2008), p. P07018.

[57] L. Ó Náraigh and J.-L. Thiffeault, *Bounds on the mixing enhancement for a stirred binary fluid*, Physica D, 237 (2008), pp. 2673–2684.

[58] A. K. O'Rourke, *Optimal mixing on the sphere*, in Proceedings of the 2010 Summer Program in Geophysical Fluid Dynamics, Woods Hole, MA, 2010, Woods Hole Oceanographic Institute.

[59] J. M. Ottino, *The Kinematics of Mixing: Stretching, Chaos, and Transport*, Cambridge University Press, Cambridge, U.K., 1989.

[60] ——, *Mixing, chaotic advection, and turbulence*, Annu. Rev. Fluid Mech., 22 (1990), pp. 207–253.

[61] G. C. Papanicolaou and S. R. S. Varadhan, *Boundary value problems with rapidly oscillating random coefficients*, in Random Fields: Rigorous Results in Statistical Mechanics and Quantum Field Theory, vol. 2 of Colloquia Mathematica Societatis János Bolyai, Amsterdam, 1979, North-Holland.

[62] R. T. Pierrehumbert, *Tracer microstructure in the large-eddy dominated regime*, Chaos Solitons Fractals, 4 (1994), pp. 1091–1110.

[63] A. Pikovsky and O. Popovych, *Persistent patterns in deterministic mixing flows*, Europhys. Lett., 61 (2003), pp. 625–631.

[64] S. Plasting and W. R. Young, *A bound on scalar variance for the advection–diffusion equation*, J. Fluid Mech., 552 (2006), pp. 289–298.

[65] O. Popovych, A. Pikovsky, and B. Eckhardt, *Abnormal mixing of passive scalars in chaotic flows*, Phys. Rev. E, 75 (2007), p. 036308.

[66] D. Rothstein, E. Henry, and J. P. Gollub, *Persistent patterns in transient chaotic fluid mixing*, Nature, 401 (1999), pp. 770–772.

[67] B. L. Sawford, *Turbulent relative dispersion*, Annu. Rev. Fluid Mech., 33 (2001), pp. 289–317.

[68] B. L. Sawford and J. C. R. Hunt, *Effects of turbulence structure, molecular diffusion and source size on scalar fluctuations in homogeneous turbulence*, J. Fluid Mech., 165 (1986), pp. 373–400.

[69] A. Sharma and N. Gupte, *Control methods for problems of mixing and coherence in chaotic maps and flows*, Pramana – Journal of Physics, 48 (1997), pp. 231–248.

[70] T. A. Shaw, *Bounds on multiscale mixing efficiencies*, in Proceedings of the 2005 Summer Program in Geophysical Fluid Dynamics, Woods Hole, MA, 2005, Woods Hole Oceanographic Institute, p. 291.

[71] T. A. Shaw, J.-L. Thiffeault, and C. R. Doering, *Stirring up trouble: Multi-scale mixing measures for steady scalar sources*, Physica D, 231 (2007), pp. 143–164.

[72] B. I. Shraiman and E. D. Siggia, *Lagrangian path integrals and fluctuations in random flow*, Phys. Rev. E, 49 (1994), pp. 2912–2927.

[73] M. Slemrod, *Stabilization of bilinear control systems with applications to non-conservative problems in elasticity*, SIAM J. Control Optim., 16 (1978), pp. 131–141.

[74] Z. B. Stone and H. A. Stone, *Imaging and quantifying mixing in a model droplet micromixer*, Phys. Fluids, 17 (2005), p. 063103.

[75] M. A. Stremler, *Mixing measures*, in Encyclopedia of Microfluidics and Nanofluidics, D. Li, ed., Springer, New York, 2008, pp. 1376–1382.

[76] M. A. Stremler and B. A. Cola, *A maximum entropy approach to optimal mixing in a pulsed source-sink flow*, Phys. Fluids, 18 (2006), p. 011701.

[77] J. Sukhatme and R. T. Pierrehumbert, *Decay of passive scalars under the action of single scale smooth velocity fields in bounded two-dimensional domains: From non-self-similar probability distribution functions to self-similar eigenmodes*, Phys. Rev. E, 66 (2002), p. 056032.

[78] G. I. Taylor, *Dispersion of soluble matter in solvent flowing slowly through a tube*, Proc. R. Soc. Lond. A, 219 (1953), pp. 186–203.

[79] J.-L. Thiffeault, *Scalar decay in chaotic mixing*, in Transport and Mixing in Geophysical Flows, J. B. Weiss and A. Provenzale, eds., vol. 744 of Lecture Notes in Physics, Berlin, 2008, Springer, pp. 3–35.

[80] J.-L. Thiffeault and S. Childress, *Chaotic mixing in a torus map*, Chaos, 13 (2003), pp. 502–507.

[81] J.-L. Thiffeault and C. R. Doering, *The mixing efficiency of open flows*, Physica D, 240 (2011), pp. 180–185.

[82] J.-L. Thiffeault, C. R. Doering, and J. D. Gibbon, *A bound on mixing efficiency for the advection–diffusion equation*, J. Fluid Mech., 521 (2004), pp. 105–114.

[83] J.-L. Thiffeault and M. D. Finn, *Topology, braids, and mixing in fluids*, Phil. Trans. R. Soc. Lond. A, 364 (2006), pp. 3251–3266.

[84] J.-L. Thiffeault and G. A. Pavliotis, *Optimizing the source distribution in fluid mixing*, Physica D, 237 (2008), pp. 918–929.

[85] D. J. Thomson, *A stochastic model for the motion of particle pairs in isotropic high-Reynolds number turbulence, and its application to the problem of concentration variance*, J. Fluid Mech., 210 (1990), pp. 113–153.

[86] A. A. Townsend, *The diffusion of heat spots in isotropic turbulence*, Proc. R. Soc. Lond. A, 209

(1951), pp. 418–430.

[87] ——, *The diffusion behind a line source in homogeneous turbulence*, Proc. R. Soc. Lond. A, 224 (1954), pp. 487–512.

[88] Y.-K. Tsang, T. M. Antonsen, Jr., and E. Ott, *Exponential decay of chaotically advected passive scalars in the zero diffusivity limit*, Phys. Rev. E, 71 (2005), p. 066301.

[89] M. R. Turner, A. D. Gilbert, and J. Thuburn, *Effective diffusion of scalar fields in a chaotic flow*, Phys. Fluids, 20 (2008), p. 107103.

[90] M. R. Turner, J. Thuburn, and A. D. Gilbert, *The influence of periodic islands in the flow on a scalar tracer in the presence of a steady source*, Phys. Fluids, 21 (2009), p. 067103.

[91] A. Vikhansky, *Enhancement of laminar mixing by optimal control methods*, Chem. Eng. Sci., 57 (2002), pp. 2719–2725.

[92] P. Welander, *Studies on the general development of motion in a two-dimensional, ideal fluid*, Tellus, 7 (1955), pp. 141–156.

[93] A. Wonhas and J. C. Vassilicos, *Mixing in fully chaotic flows*, Phys. Rev. E, 66 (2002), p. 051205.